GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 4, 2002, 16:08:57 ; Search time 165.17 Seconds
(without alignments)
316.066 Million cell updates/sec

Title: US-09-052-089A-2
Perfect score: 2393
Sequence: 1 MPILSLCTICSDFFDHSRDV.....VRIKTVSSASQPKLDTFLCQ 470

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1826.5	76.3	469	20	AAY30149	Amino acid sequenc
2	1819.5	76.0	469	19	AAW37881	BRCA1 modulator pr
3	281.5	11.8	455	22	ABB61289	Drosophila melanog
4	195.5	8.2	1690	22	ABB61144	Drosophila melanog
5	195.5	8.2	1690	22	ABB61173	Drosophila melanog
6	192.5	8.0	1325	18	AAW19540	Male-enhanced anti
7	192.5	8.0	1325	20	AAW94391	Mouse male enhance
8	190	7.9	2056	22	ABB59344	Drosophila melanog
9	188.5	7.9	574	22	AAB95497	Human protein sequ
10	179.5	7.5	1988	22	AAM40999	Human polypeptide
11	179.5	7.5	1988	22	AAM41000	Human polypeptide
12	178	7.4	482	22	ABB71396	Drosophila melanog
13	177	7.4	1851	22	ABG01723	Novel human diagno
14	177	7.4	1960	22	AAM78854	Human protein SEQ
15	177	7.4	2143	22	ABG01716	Novel human diagno
16	176.5	7.4	1975	22	ABB62094	Drosophila melanog
17	175.5	7.3	1017	22	AAE02246	Domestic mite Bt11
18	175	7.3	1177	22	AAB96721	Putative P. abyssi
19	174.5	7.3	875	22	AAE02245	Domestic mite Bt11
20	174.5	7.3	878	22	AAE02242	Domestic mite Bt11
21	174	7.3	1374	22	AAB69070	Human male enhance
22	174	7.3	1489	22	ABB59948	Drosophila melanog
23	173.5	7.3	1203	22	AAM79264	Human protein SEQ
24	172.5	7.2	864	22	AAM40292	Human polypeptide
25	172.5	7.2	2482	16	AAR72826	Human mitosin. Ho
26	172.5	7.2	2482	19	AAW23996	Human mitosin amin
27	172.5	7.2	3248	17	AAR99795	Kinetochore protei
28	172	7.2	717	21	AAB21231	Tomato LeMPF1. Ly
29	171.5	7.2	1456	22	ABB58673	Drosophila melanog
30	170.5	7.1	866	21	AAY86194	Nuclear transport
31	170.5	7.1	1752	20	AAY07031	Breast cancer asso
32	170.5	7.1	1948	22	ABG21233	Novel human diagno
33	170.5	7.1	1963	22	AAM79838	Human protein SEQ
34	170.5	7.1	2918	22	ABG27218	Novel human diagno
35	170	7.1	1093	14	AAR42818	TMF. Homo sapiens
36	170	7.1	2017	22	ABG06301	Novel human diagno
37	169.5	7.1	561	19	AAW63043	Streptococcus uber
38	169	7.1	1090	21	AAY59270	Human huntingtin-i
39	168.5	7.0	560	22	AAU31067	Novel human secret
40	168.5	7.0	576	16	AAR66929	AMML chromosome in
41	168.5	7.0	731	22	AAB48573	Human breast cance
42	168.5	7.0	816	16	AAR66931	AMML chromosome in
43	168.5	7.0	885	16	AAR66930	AMML chromosome in
44	168.5	7.0	2400	22	ABG20278	Novel human diagno
45	168.5	7.0	2415	22	ABG20279	Novel human diagno

ALIGNMENTS

RESULT 1
AAY30149
ID AAY30149 standard; Protein; 469 AA.
XX
AC AAY30149;
XX
DT 27-OCT-1999 (first entry)
XX
DE Amino acid sequence of a BRCA1 modulator protein.
XX
KW Modulator protein; BRCA1; tumour suppressor protein; breast cancer;
KW ovarian cancer; cell growth; cell proliferation.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 3..32
FT /note= "zinc finger motif"
FT Region 230..255
FT /note= "leucine zipper motif"
XX
PN US5948643-A.
XX
PD 07-SEP-1999.
XX
PF 13-AUG-1997; 97US-0968751.
XX
PR 13-AUG-1997; 97US-0968751.
XX
PA (ONYX-) ONYX PHARM INC.
XX
PI Lingenfelter C, Polakis PG, Rubinfeld B, Vuong TT;
XX
DR WPI; 1999-517952/43.

```
DR   N-PSDB; AAX86754.
XX
PT   Modulator proteins that bind to and modulate the activity of the
PT   BRCA1 tumour suppressor gene product, useful for the treatment of
PT   ovarian and breast cancer
XX
PS   Example 1; Fig 1; 35pp; English.
XX
CC   The present sequence represents a modulator protein, that binds to and
CC   modulate the activity of the BRCA1 gene product (BRCA1). The BRCA1
CC   protein has been characterized as a tumour suppressor protein.
CC   Alterations in the amino acid sequence of BRCA1 causes breast and ovarian
CC   cancers by removing the controls on cell growth and proliferation.
CC   Research has shown that different regions on the BRCA1 molecule have
CC   different effects on cell growth and tumour suppression (e.g. full length
CC   truncated BRCA1 has no effect on breast cancer cell growth but will
CC   inhibit ovarian cancer cell growth). It has been suggested that different
CC   host cell factors (e.g. proteins) interact with different regions of the
CC   BRCA1 to control its function. The identification of these proteins
CC   (e.g. BRCA1MP) will facilitate the development of novel diagnostic
CC   methods and new therapeutics for identifying and treating cancers caused
CC   by changes in the expression or activity of BRCA1.
XX
SQ   Sequence   469 AA;


  Query Match             76.3%;  Score 1826.5;  DB 20;  Length 469;
  Best Local Similarity   77.6%;  Pred. No. 1.4e-147;
  Matches  363;  Conservative   41;  Mismatches   63;  Indels    1;  Gaps    1;

Qy       1 MPILSLCTICSDFFDHSRDVAAIHCGHTFHLQCLIQWFETAPSRTCPQCRIQVGKKTIIN 60
           ||| :|||||||||||||||||||||||||||||||||||||||||||||||||||:||||
Db       1 mpiralcticsdffdhsrdvaaihcghtfhlqcliqwfetapsrtcpqcriqvgkrtiin 60

Qy      61 KLFFDLAQEEENVLDAEFLKNELDSVKAQLSQKDREKRDSQAIIDTLRDTLEERNATVES 120
           |||||||||||||||||||||||:|:|||||||:|||||| |||||||||||||||| |
Db      61 klffdlaqeeenvldaeflkneldnvraqlsqkdekrdsqviidtlrdtleernatvvs 120

Qy     121 LQNALNKAEMLCSTLKKQMKFLEQRQDETKQAREEAHRLKCKMKTMEQIELLLQSQRSEV 180
           || || |||||||||||||:|||:|||||||:||| ||: ||||||||||||||| ||
Db     121 lqqalgkaemlcstlkkqmkyleqqqdetkqaqeearrlrskmktmeqielllqsqrpev 180

Qy     181 EEMIRDMGVGQSAVEQLAVYCVSLKKEYENLKEARKATGELADRLKKDLVSSRSKLKTLN 240
           ||||||||||||||||||||||||||||||||||:||:||:|:||| |||||||:|:
Db     181 eemirdmgvgqsaveqlavycvslkkeyenlkearkasgevadklrkdlfssrsklqtvy 240

Qy     241 TELDQAKLELRSAQKDLQSADQEITSLRKKSDDPPGNLEPASATNETVSRLVFESPAPVE 300
           :|||||||||:|||||||||||:|| ||:||       |      :||| ||| |||||||
Db     241 seldqaklelksaqkdlqsadkeimslkkkltmlqetlnlppvasetvdrlvlespapve 300

Qy     301 MMNPRLHQPPFGDEIDLNTTFDVNTPPTQTSGSQHCLPKKLCLERARSPMQNVLKKVHKV 360
            :| :| :| | |:|||| ||||:||| : | |||   :|||||:: ||:|:| ||: |
Db     301 -vnlklrrpsfrddidlnatfdvdtpparpsssqhgyyeklclekshspiqdvpkkickg 359

Qy     361 SKPESQLSLGGQRCVGELDEELAGAFPLFIRNAVLGQKQPNRTTAESRSSTDVVRIGFDG 420
            : |||||||| | || |||| ||||:|:|||:|||||| |  :||  | |||| ||||
Db     360 prkesqlslggqscagepdeelvgafpifvrnailgqkqpkrprsesscskdvvrtgfdg 419

Qy     421 LGGRTKFIQPRDTTIIRPVPVKSKAKSKQKVRIKTVSSASQPKLDTFL 468
           |||||||||| || :|||:||| | | ||:||:||| |  | ||||||
Db     420 lggrtkfiqptdtvmirplpvkpktkvkqrvrvktvpslfqakldtfl 467


RESULT   2
AAW37881
ID   AAW37881 standard; Protein; 469 AA.
XX
AC   AAW37881;
XX
DT   28-AUG-1998  (first entry)
XX
DE   BRCA1 modulator protein 091-21A31.
XX
KW   BRCA1 modulator protein; 091-21A31; breast cancer antigen 1;
KW   tumour suppressor protein; diagnosis; therapy; human.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          3..54
FT                   /note= "zinc finger motif"
FT   Domain          229..255
FT                   /note= "leucine zipper motif"
XX
PN   WO9810066-A1.
XX
PD   12-MAR-1998.
XX
PF   06-AUG-1997;   97WO-US13944.
XX
PR   04-SEP-1996;   96US-0025601.
XX
PA   (ONYX-) ONYX PHARM INC.
XX
PI   Ligenfelter C,  Polakis P,  Rubinfeld B,  Vuong TT;
XX
DR   WPI; 1998-193616/17.
DR   N-PSDB; AAV29062.
XX
PT   Breast cancer antigen 1 modulator protein - useful for diagnosing
PT   diseases involving unwanted cell growth, e.g. breast cancer, and for
PT   producing therapeutics for treatment of such diseases
XX
PS   Example 1; Fig 1; 73pp; English.
XX
CC   This polypeptide comprises a 53 kDa BRCA1 modulator protein that
CC   binds to the tumour suppressor gene product BRC1, and which is
CC   characterised by a zinc finger domain and a leucine zipper motif.
CC   Its amino acid sequence was deduced from the nucleotide sequence
CC   of a cDNA clone (see AAV29062), designated 091-21A31 (ATCC 98141),
CC   isolated from a HeLa cell cDNA library using a yeast two-hybrid
CC   assay.  3 cDNA clones (see also AAV29063-64) coding for BRCA1
CC   modulator proteins (see AAW37881-83) have been characterised. Vectors
CC   and host cells comprising the isolated nucleic acid sequences are
CC   claimed, as well as a process for producing BRCA1 modulator protein
CC   by culturing these host cells.  BRCA1 modulator proteins and nucleic
CC   acids can be used to diagnose diseases involving unwanted cell
CC   growth, e.g. breast cancer, and to identify compounds that alter
CC   BRCA1 interaction with BRCA1 modulators for the treatment of such
CC   diseases.
XX
SQ   Sequence   469 AA;


  Query Match             76.0%;  Score 1819.5;  DB 19;  Length 469;
  Best Local Similarity   77.4%;  Pred. No. 5.6e-147;
  Matches  362;  Conservative   41;  Mismatches   64;  Indels    1;  Gaps    1;

Qy       1 MPILSLCTICSDFFDHSRDVAAIHCGHTFHLQCLIQWFETAPSRTCPQCRIQVGKKTIIN 60
           ||| :|||||||||||||||||||||||||||||||||||||||||||||||||||:||||
Db       1 mpiralcticsdffdhsrdvaaihcghtfhlqcliqwfetapsrtcpqcriqvgkrtiin 60

Qy      61 KLFFDLAQEEENVLDAEFLKNELDSVKAQLSQKDREKRDSQAIIDTLRDTLEERNATVES 120
           |||||||||||||||||||||||:|:|||||||:|||||| |||||||||||||||| |
Db      61 klffdlaqeeenvldaeflkneldnvraqlsqkdekrdsqviidtlrdtleernatvvs 120

Qy     121 LQNALNKAEMLCSTLKKQMKFLEQRQDETKQAREEAHRLKCKMKTMEQIELLLQSQRSEV 180
           || || |||||||||||||:|||:|||||||:||| ||: ||||||||||||||| ||
Db     121 lqqalgkaemlcstlkkqmkyleqqqdetkqaqeearrlrskmktmeqielllqsqrpev 180

Qy     181 EEMIRDMGVGQSAVEQLAVYCVSLKKEYENLKEARKATGELADRLKKDLVSSRSKLKTLN 240
           ||||||||||||||||||||||||||||||||||:||:||:|:||| |||||||:|:
Db     181 eemirdmgvgqsaveqlavycvslkkeyenlkearkasgevadklrkdlfssrsklqtvy 240

Qy     241 TELDQAKLELRSAQKDLQSADQEITSLRKKSDDPPGNLEPASATNETVSRLVFESPAPVE 300
           :|||||||||:|||||||||||:|| ||:||       |      :||| ||| |||||||
Db     241 seldqaklelksaqkdlqsadkeimslkkkltmlqetlnlppvasetvdrlvlespapve 300
```

```
Db    241 seldqaklelksaqkdlqsadkeimslkkkltmlqetlnlppvasetvdrlvlespapve 300

Qy    301 MMNPRLHQPPFGDEIDLNTTFDVNTPPTQTSGSQHCLPKKLCLERARSPMQNVLKKVHKV 360
Db    301 -vnlklrrpsfrddidlnatfdvdtpparpsssqhgyyeklclekshspiqdvpkkickg 359

Qy    361 SKPESQLSLGGQRCVGELDEELAGAFPLFIRNAVLGQKQPNRTTAESRSSTDVVRIGFDG 420
Db    360 prkesqlslggqscagepdeelvgafpifvrnailgqkqpkrprsesscskdvvrtgfdg 419

Qy    421 LGGRTKFIQPRDTTIIRPVPVKSKAKSKQKVRIKTVSSASQPKLDTFL 468
Db    420 lggrtkfiqptdtvmirplpvkpktkvkqrvrvktvpslfqaklhtfl 467


RESULT   3
ABB61289
ID   ABB61289 standard; Protein; 455 AA.
XX
AC   ABB61289;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 10659.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL05392.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 10659; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   455 AA;


  Query Match           11.8%;  Score 281.5;  DB 22;  Length 455;
  Best Local Similarity   22.0%;  Pred. No. 1.7e-15;
  Matches  106;  Conservative   99;  Mismatches  197;  Indels   79;  Gaps   17;

Qy      7 CTICSDFFDHSRDVAAIHCGHTFHLQCLIQWFETAPSRTCPQCRIQVGKKTIINKLFFDL 66

Db      6 cvicaelfgqadevfatvcghmfhhnclnqwldr--sktcpqcrnkcttrnif-rvyfnl 62

Qy     67 AQEEENVLDAEFLKNELDSVKAQLSQKDREKRDSQAIIDTLRDTLEERNATVESLQNALN 126
Db     63 anldvshidvgslqeqldnamlsmkmvekernkdeqqirdlketqkkclktiagleqkvq 122

Qy    127 KAEMLCSTLKKQMKFLEQRQDETKQAREEAHRLKCKMKTMEQIELLLQSQRSEVEEMIRD 186
Db    123 kkdflissyveqigvlksdahvvdglrkenktlksqiqsmegisailaagsadadrllkn 182

Qy    187 MGVGQSAVEQLAVYCVSLKKEYENLK----EARKATGELADRLKKDLVSSRSKLKTLNT- 241
Db    183 ----eadphvlanwvstlkrelrqceskktelrnvvkvvqndlrkeielkryahnipnvv 238

Qy    242 ------ELDQAKLELRSA--QKDLQSADQEITSLRKKS---DDPPGNLEPASATNETVSR 290
Db    239 fiidmcsfdcskleervshlesdlyqaqeklqafenktayldsp----nascglnsnila 294

Qy    291 LVFE------SPAPVEMMNPRLHQPPFGDEIDLNTTFDVNTPPTQTSGSQHCLPKK--LC 342
Db    295 lkreerrttisptvkenik----------rieestspylni-ksssvglahllntkgnig 343

Qy    343 LERAR-SPMQNVLKKVHKVSKPESQLSLGGQRCVGELDEELAGAFPLFIRNAVLGQKQPN 401
Db    344 lakskispikgvgggvsmts----------gtirktssdlsekysif--------kkpr 384

Qy    402 RTTAESRSSTDVVRIG----FDGLGGRTKFIQPRDTTIIRPVPVKSKAKSKQKVRIKTVS 457
Db    385 lllgsssssaltattgsnfvyngmggsek---------vdpfaqraeeeglstirsstal 435

Qy    458 S 458
Db    436 s 436


RESULT   4
ABB61144
ID   ABB61144 standard; Protein; 1690 AA.
XX
AC   ABB61144;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 10224.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL05247.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 10224; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
```

```
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1690 AA;


  Query Match             8.2%;  Score 195.5;  DB 22;  Length 1690;
  Best Local Similarity   22.7%;  Pred. No. 2.5e-07;
  Matches  110;  Conservative   74;  Mismatches  163;  Indels  137;  Gaps   20;

Qy      50 RIQVGKKTIINKLFFDLAQEEENVLDAEFLKNE------LDSVKAQLSQKDREKRDSQAI 103
           :||: |::|  :|       ::| |: :| | :      |  :||| :||| |  :|
Db     724 qiqlekesieqqlal-----kqnele-dfqkkqsesevhlqeikaqntqkdfelvesges 777

Qy     104 IDTLRDTLEERNATVESLQNALNKAEMLCSTLKKQMK-FLEQRQDETKQAREEAHRLKCK 162
           :  |:  ||::    | || || :       |||: :  ::::: | :| : ::   :
Db     778 lkklqqqleqktlgheklqaalee-------lkkeketiikekeqelqqlqsksaesesa 830

Qy     163 MKTME-QIELLLQSQRSEVEEMIRDMGVGQSAVEQLAVYC----VSLKKEYENL----KE 213
           :| :: |:| | |   :  ||  : :     : ||          ||    ||    |:
Db     831 lkvvqvqleqlqqqaaasgeegsktvaklhdeisqlksqaeetqselkstqsnleakskq 890

Qy     214 ARKATGELADRLKKD--LVSSRSKLKTLNTELDQAKLELRSAQKDLQSADQEITSLRKKS 271
             | | :  ||   |:   :|||   :|: : :  | |   |::|  ::
Db     891 leaangsleeeakksghlleqitklk---sevgetqaalsschtdvesktkq-------- 939

Qy     272 DDPPGNLEPASATNETVSRLVFESPA-------PVEMMNPRLHQ---------------- 308
                || |:|  | |::   || |        |: :   ||
Db     940 ------leaanaalekvnkeyaesraeasdlqdkvkeitdtlhaelqaerssssalhtkl 993

Qy     309 PPFGDEI-----DLNTTFDVNTPPTQTSGSQHCLPKKLCLERARSPMQ------------ 351
             | |||     :| :  |        : ||  | |:  |:  |  :|
Db     994 skfsdeiatghkeltskad--------awsqemlqkekelqelrqqlqdsqdsqtklkae 1045

Qy     352 ------------NVLKKVHKVSKPESQLSLGGQRCVGELDEELAGAFPLFIRNAVLGQK 398
                       |: ::| |      :|| | |  : :| | |       | || |  |
Db    1046 gerkeksfeesiknlqeevtkaktenlelstgtqttikdlqerle------itnaelqhk 1099

Qy     399 QPNRTTAESRSSTDVVRIGFDGLGGRTKFIQPRD-TTIIRPVPVKSKAKSKQKVRIKTVS 457
                |  :| |  :|               |  |::  : | :   |     : ||
Db    1100 -------ekmasedaqkia-------------dlktlveaiqvananisatnaelstvl 1138

Qy     458 SASQ 461
              |
Db    1139 evlq 1142


RESULT   5
ABB61173
ID   ABB61173 standard; Protein; 1690 AA.
XX
AC   ABB61173;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 10311.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL05276.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 10311; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1690 AA;


  Query Match             8.2%;  Score 195.5;  DB 22;  Length 1690;
  Best Local Similarity   22.7%;  Pred. No. 2.5e-07;
  Matches  110;  Conservative   74;  Mismatches  163;  Indels  137;  Gaps   20;

Qy      50 RIQVGKKTIINKLFFDLAQEEENVLDAEFLKNE------LDSVKAQLSQKDREKRDSQAI 103
           :||: |::|  :|       ::| |: :| | :      |  :||| :||| |  :|
Db     724 qiqlekesieqqlal-----kqnele-dfqkkqsesevhlqeikaqntqkdfelvesges 777

Qy     104 IDTLRDTLEERNATVESLQNALNKAEMLCSTLKKQMK-FLEQRQDETKQAREEAHRLKCK 162
           :  |:  ||::    | || || :       |||: :  ::::: | :| : ::   :
Db     778 lkklqqqleqktlgheklqaalee-------lkkeketiikekeqelqqlqsksaesesa 830

Qy     163 MKTME-QIELLLQSQRSEVEEMIRDMGVGQSAVEQLAVYC----VSLKKEYENL----KE 213
           :| :: |:| | |   :  ||  : :     : ||          ||    ||    |:
Db     831 lkvvqvqleqlqqqaaasgeegsktvaklhdeisqlksqaeetqselkstqsnleakskq 890

Qy     214 ARKATGELADRLKKD--LVSSRSKLKTLNTELDQAKLELRSAQKDLQSADQEITSLRKKS 271
             | | :  ||   |:   :|||   :|: : :  | |   |::|  ::
Db     891 leaangsleeeakksghlleqitklk---sevgetqaalsschtdvesktkq-------- 939

Qy     272 DDPPGNLEPASATNETVSRLVFESPA-------PVEMMNPRLHQ---------------- 308
                || |:|  | |::   || |        |: :   ||
Db     940 ------leaanaalekvnkeyaesraeasdlqdkvkeitdtlhaelqaerssssalhtkl 993

Qy     309 PPFGDEI-----DLNTTFDVNTPPTQTSGSQHCLPKKLCLERARSPMQ------------ 351
             | |||     :| :  |        : ||  | |:  |:  |  :|
Db     994 skfsdeiatghkeltskad--------awsqemlqkekelqelrqqlqdsqdsqtklkae 1045

Qy     352 ------------NVLKKVHKVSKPESQLSLGGQRCVGELDEELAGAFPLFIRNAVLGQK 398
                       |: ::| |      :|| | |  : :| | |       | || |  |
Db    1046 gerkeksfeesiknlqeevtkaktenlelstgtqttikdlqerle------itnaelqhk 1099

Qy     399 QPNRTTAESRSSTDVVRIGFDGLGGRTKFIQPRD-TTIIRPVPVKSKAKSKQKVRIKTVS 457
                |  :| |  :|               |  |::  : | :   |     : ||
Db    1100 -------ekmasedaqkia-------------dlktlveaiqvananisatnaelstvl 1138

Qy     458 SASQ 461
              |
Db    1139 evlq 1142
```

```
RESULT   6
AAW19540
ID   AAW19540 standard; Protein; 1325 AA.
XX
AC   AAW19540;
XX
DT   16-SEP-1997  (first entry)
XX
DE   Male-enhanced antigen-2.
XX
KW   Mouse; MEA-2; detecting mutation.
XX
OS   Mus musculus domesticus.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 305..320
FT                   /note= "Not shown in the specification"
XX
PN   JP09121869-A.
XX
PD   13-MAY-1997.
XX
PF   07-NOV-1995;   95JP-0311638.
XX
PR   07-NOV-1995;   95JP-0311638.
XX
PA   (ITOH-) ITO HAM KK.
XX
DR   WPI; 1997-314229/29.
DR   N-PSDB; AAT74034.
XX
PT   Male-enhanced antigen Mea-2 gene - especially from mouse, useful for
PT   detecting mutation(s)
XX
PS   Claim 8; Page 9-10; 13pp; Japanese.
XX
CC   The present sequence represents male-enhanced antigen-2 (MEA-2), which
CC   has been derived from a domestic mouse. The polynucleotide encoding
CC   the protein can be used for the detection of mutations affecting the
CC   MEA-2 gene.
XX
SQ   Sequence   1325 AA;

  Query Match             8.0%;  Score 192.5;  DB 18;  Length 1325;
  Best Local Similarity   22.4%;  Pred. No. 3.2e-07;
  Matches   94;  Conservative   80;  Mismatches  156;  Indels   89;  Gaps   16;

Qy      65 DLAQEEENVLDA-EFLKNE-------LDSVKAQLSQKDREKR-------DSQAIIDTLRD 109
           :| :| ::  ||  ||:||       | | |:   : ||  |       ::  ::: ||
Db     590 elqreadsredaihflqnekivlevalqsaksdkeeldrgarrleedteetsglleqlrq 649

Qy     110 TLEERNATVESLQNALNKAEMLCSTLKKQM-----KFLEQ-------RQDET------KQ 151
           |  ::  || ||          :||:|||     :|::|       |:| |       :
Db     650 dlavksnqvehlqqe-------tatlrkqmqkvkeqfvqqkvmveayrrdatskdqline 702

Qy     152 AREEAHRLKCKMKTMEQIELLLQSQRSEVE----EMIRDMGVGQSAVEQLAVYCVSLKKE 207
           :    ||  :|| : |  : || ::  ||     : :|| :    : :|  :  |::||
Db     703 lkatkkrldsemkelrqeliklqgekktvevehsrlqkdmslvhqqmaeleghlqsvqke 762

Qy     208 YEN-----------------LKEA----RKATGELADRLKKDLVSSRSKLKTLNTELDQA 246
           :                   | ||    :|   ||    || :   : |:| | ::|  |
Db     763 rdemeihlqslkfdkeqmialteanetlkkqieelqqeakkaiteqkqkmkrlgsdltsa 822

Qy     247 KLELRSAQKDLQSADQEITSLRKKSDDPPGNLEPASATNETVSRLVFESPAPVEMMNPRL 306
           : |:::  |  ::|   :: | ::  :    |    ||:  :::|  :|       :| |
Db     823 qkemktkhkayena---vsilsrrlqea---laskeatdaelnqlraqstg--gssdpvl 874

Qy     307 HQPPFGDEIDLNTT--------FDVNTPPTQTSGS-QHCLPKKLCLERARSPMQNVLKKV 357
           |:     |::|            ::    | ||   :    | | ||      :   :|:
Db     875 hekiralevelqnvgqskillekelqevitmtsqeleesrekvleledelqesrgfrrki 934

Qy     358 HKVSKPESQLSLGGQRCVGELDEELAGAFPLFIRNAVLGQKQPNRTTAESRSSTDVVRI 416
           :: :   :|:|       ||: |      |   || | :      || ::   |:|::
Db     935 krleesnkklal-------elehergkltglgqsnaalrehnsiletalakreadlvql 986


RESULT   7
AAW94391
ID   AAW94391 standard; Protein; 1325 AA.
XX
AC   AAW94391;
XX
DT   14-APR-1999  (first entry)
XX
DE   Mouse male enhanced antigen 2.
XX
KW   Mouse; male enhanced antigen 2; Mea-2; Mus musculus domesticus;
KW   spermatogenesis; regulation; contraceptive; sterile; inhibition.
XX
OS   Mus sp.
XX
PN   JP11018622-A.
XX
PD   26-JAN-1999.
XX
PF   04-JUL-1997;   97JP-0179490.
XX
PR   04-JUL-1997;   97JP-0179490.
XX
PA   (ITOH-) ITO HAM KK.
XX
DR   WPI; 1999-160962/14.
DR   N-PSDB; AAX04132.
XX
PT   Regulation of spermatogenesis using Mea-2 gene information  - using
PT   anti-sense oligo- or poly:nucleotide(s), used for production of
PT   contraceptives
XX
PS   Claim 4; Page 8-12; 27pp; Japanese.
XX
CC   The present sequence represents mouse male enhanced antigen 2 (Mea-2).
CC   The present invention describes the regulation of spermatogenesis by
CC   using Mea-2 information. A non-human living organism can have its
CC   spermatogenesis inhibited by breakage of the whole or part of the Mea-2
CC   gene. Also described are: (1) the creation of the spermatogenesis-
CC   inhibited organism; (2) a drug composition containing an oligonucleotide
CC   or polynucleotide containing base sequences that pair with at least part
CC   of the Mea-2 gene and are able to inhibit the expression of Mea-2 gene;
CC   and (3) the creation of an aimed gene-possessing organism using the
CC   spermatogenesis inhibited organism. The organism is useful for producing
CC   contraceptive drugs.
XX
SQ   Sequence   1325 AA;

  Query Match             8.0%;  Score 192.5;  DB 20;  Length 1325;
  Best Local Similarity   22.4%;  Pred. No. 3.2e-07;
  Matches   94;  Conservative   80;  Mismatches  156;  Indels   89;  Gaps   16;

Qy      65 DLAQEEENVLDA-EFLKNE-------LDSVKAQLSQKDREKR-------DSQAIIDTLRD 109
           :| :| ::  ||  ||:||       | | |:   : ||  |       ::  ::: ||
Db     590 elqreadsredaihflqnekivlevalqsaksdkeeldrgarrleedteetsglleqlrq 649

Qy     110 TLEERNATVESLQNALNKAEMLCSTLKKQM-----KFLEQ-------RQDET------KQ 151
           |  ::  || ||          :||:|||     :|::|       |:| |       :
Db     650 dlavksnqvehlqqe-------tatlrkqmqkvkeqfvqqkvmveayrrdatskdqline 702

Qy     152 AREEAHRLKCKMKTMEQIELLLQSQRSEVE----EMIRDMGVGQSAVEQLAVYCVSLKKE 207
           :    ||  :|| : |  : || ::  ||     : :|| :    : :|  :  |::||
Db     703 lkatkkrldsemkelrqeliklqgekktvevehsrlqkdmslvhqqmaeleghlqsvqke 762

Qy     208 YEN-----------------LKEA----RKATGELADRLKKDLVSSRSKLKTLNTELDQA 246
           :                   | ||    :|   ||    || :   : |:| | ::|  |
```

```
Db    763 rdemeihlqslkfdkeqmialteanetlkkqieelqqeakkaiteqkqkmkrlgsdltsa 822

Qy    247 KLELRSAQKDLQSADQEITSLRKKSDDPPGNLEPASATNETVSRLVFESPAPVEMMNPRL 306
Db    823 qkemktkhkayena---vsilsrrlqea---laskeatdaelnqlraqstg--gssdpvl 874

Qy    307 HQPPFGDEIDLNTT--------FDVNTPPTQTSGS-QHCLPKKLCLERARSPMQNVLKKV 357
Db    875 hekiralevelqnvgqskillekelqevitmtsqeleesrekvleledelqesrgfrrki 934

Qy    358 HKVSKPESQLSLGGQRCVGELDEELAGAFPLFIRNAVLGQKQPNRTTAESRSSTDVVRI 416
Db    935 krleesnkklal-------elehergkltglgqsnaalrehnsiletalakreadlvql 986


RESULT   8
ABB59344
ID   ABB59344 standard; Protein; 2056 AA.
XX
AC   ABB59344;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 4824.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL03447.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 4824; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   2056 AA;


  Query Match             7.9%;  Score 190;  DB 22;  Length 2056;
  Best Local Similarity   23.8%;  Pred. No. 9.9e-07;
  Matches   88;  Conservative   61;  Mismatches  145;  Indels   76;  Gaps   14;

Qy     56 KTIINKLFFDLAQEEENVLDAE--FLKNELDSVKAQLSQKDREKRDSQ----------AI 103

Db   1302 ktvlek--------akgtleaenadlatelrsvnssrqendrrrkqaesqiaelqvklae 1353

Qy    104 IDTLRDTLEER----NATVESLQNALNKAEMLCSTLKKQMKFLEQRQDETKQAREEAHR- 158
Db   1354 ierarselqekctklqqeaenitnqleeaelkasaavksasnmesqlteaqqlleeetrq 1413

Qy    159 ---LKCKMKTMEQIELLLQSQRSEVEEMIRDMGVGQSAVEQLAVYCVSLKKEYENLKEAR 215
Db   1414 klglssklrqiesekealqeqleeddeakrny---erklaevttqmqeikkkaeedadla 1470

Qy    216 KATGELADRLKKDLVSSRSKLKTL---NTELDQAKLELRSAQKD----LQSADQEITSLR 268
Db   1471 keleegkkrlnkdiealerqvkeliaqndrldkskkkiqseledatieleaqrtkvlele 1530

Qy    269 KKSDDPPGNLEPASATNETVSRLVFESPAPVEMMNPRLHQPPFGDEIDLNTTFDVNTPPT 328
Db   1531 kkqk----nfdkilaeekaiseqiaqerdtaereareketkvlsvsreldeafd------ 1580

Qy    329 QTSGSQHCLPKKLCLERARSPMQNVL----------KKVHKVSKP----ESQLS-LGGQ 372
Db   1581 ----------kiedlenkrktlqnelddlantqgtadknvhelekakralesqlaelkaq 1630

Qy    373 RCVGELDEEL 382
Db   1631 n--eeleddl 1638


RESULT   9
AAB95497
ID   AAB95497 standard; Protein; 574 AA.
XX
AC   AAB95497;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:18041.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 18041; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
```

```
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   574 AA;


  Query Match             7.9%;  Score 188.5;  DB 22;  Length 574;
  Best Local Similarity   20.4%;  Pred. No. 2.1e-07;
  Matches   80;  Conservative   84;  Mismatches  147;  Indels   81;  Gaps   14;

Qy      48 QCRIQVGKKTIINKL--FFDLAQEEENVLDAEFLKNELDSVKAQLSQKDREKRDSQAIID 105
           | : :   || : ||    ::||:     |  ||:||:  :: || |||||  | :  ::
Db      80 qakfenqlktkvmklenelemaqqsaggrdtrflrneicqlekqleqkdreledmekele 139

Qy     106 ---TLRDTLEERNATVESLQNAL--------NKAEMLCSTLKKQMKFLEQRQDE--TKQA 152
               : : |  ||   |:  : |         | | ||  :    | :: :::   :::
Db     140 kekkvneqlalrneeaenensklrrenkrlkkkneqlcqdiidyqkqidsqketllsrrg 199

Qy     153 REEAHRLKCKMKTM----------------EQIELLLQSQRSEVEEMIRDMGVGQSAVEQ 196
            :  :| :   |                 |:||:  |  |  :|| :::|
Db     200 edsdyrsqlskknyeliqyldeiqtlteanekievqnqemrknleesvqem--------- 250

Qy     197 LAVYCVSLKKEYENLKEARKATGELADRLKK------------DLVSSRSK-----LKT 238
                :  ||  :|     |  : |:|||             ||: |:::     :
Db     251 -----ekmtdeynrmkaivhqtdnvidqlkkendhyqlqvqeltdllkskneeddpimva 305

Qy     239 LNTELDQAKLELRSAQKDLQSADQEITSLRKKSDDPPGNLEPASATNETVSRLVFESPAP 298
           :| :::: || | |   ::    | : :||:|  :    |:   :    : : : |  :
Db     306 vnakveewklilsskddeiieyqqmlhnlreklkn--aqldadksnvmalqqgiqerdsq 363

Qy     299 VEMMNPRLHQPPFGDEIDLNTTF--DVNTPPTQTSGSQHCLPKKLCLERARSPMQ---NV 353
           ::|:  :: |  :  |:: ||    |:     :  |:        ::    :|   ::
Db     364 ikmlteqveq--ytkemekntciiedlknelqrnkga------stlsqqthmkiqstldi 415

Qy     354 LKKVHKVSKPESQLSLGGQRCVGELDEELAGA 385
           ||:  | ::  ::|:    |   | |:||  |
Db     416 lkektkeaertaelaeadar---ekdkelvea 444


RESULT  10
AAM40999
ID   AAM40999 standard; Protein; 1988 AA.
XX
AC   AAM40999;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 5930.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   N-PSDB; AAI60155.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Example 2; SEQ ID NO 5930; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence   1988 AA;


  Query Match             7.5%;  Score 179.5;  DB 22;  Length 1988;
  Best Local Similarity   21.8%;  Pred. No. 7.4e-06;
  Matches   76;  Conservative   77;  Mismatches  140;  Indels   55;  Gaps   10;

Qy      68 QEEENVLDAEFLKNELDSVKAQLSQKDREKRDSQAIIDTLRDTLEERNATVESLQNALNK 127
           | ::   | | | :  | |   : : :: ||  :  :: :|  |||    ::: ::|  :
Db    1517 qnkqlradmedlmsskddvgknvhelekskraleqqveemrtqleeledelqatedaklr 1576

Qy     128 AEMLCSTLKKQMKFLEQRQDE---------TKQAREEAHRLKCKMK-------TMEQIEL 171
            |:    :| | :   | :|| '          || ||    |: : |       : :::|:
Db    1577 levnmqamkaqferdlqtrdeqneekkrllikqvreleaelederkqralavaskkkmei 1636

Qy     172 LLQSQRSEVE--EMIRDMGVGQSAVEQLAVYCVSLKKEYENLKEARKATGELADRLKKDL 229
            |:   :::|     ||       ::||      :|     |:|||  :  |:  : |:
Db    1637 dlkdleaqieaankard-----evikqlrklqaqmkdyqreleearasrdeifaqske-- 1689

Qy     230 VSSRSKLKTLNTELDQAKLELRSAQKDLQSADQEITSLRKKSDDPPGNLEPASATNETVS 289
             |  |||:|  |: | : || |:::  : |:||   |   :|:   :    ||  :
Db    1690 --sekklksleaeilqlqeelasserarrhaeqerdel---adeitnsasgksalldekr 1744

Qy     290 RL----------VFESPAPVEMMNPRLHQPPFGDEIDLNTTFDVNTPPTQTSGSQHCLPK 339
           ||          : |  : :|::| |  :          ||  |:|   : :  :    |
Db    1745 rleariaqleeeleeeqsnmellndrfrk----------ttlqvdtlnaelaaersaaqk 1794

Qy     340 ----KLCLERARSPMQNVLKKVHKVSKPESQLSLGG-QRCVGELDEEL 382
```

```
              :  |||    ::  |:::    | : : ::   :  :|:|:|:|
Db   1795 sdnarqqlerqnkelkaklqelegavkskfkatisaleakigqleeql 1842


RESULT 11
AAM41000
ID   AAM41000 standard; Protein; 1988 AA.
XX
AC   AAM41000;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 5931.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   N-PSDB; AAI60156.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Example 2; SEQ ID NO 5931; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence   1988 AA;


  Query Match             7.5%;  Score 179.5;  DB 22;  Length 1988;
  Best Local Similarity   21.8%;  Pred. No. 7.4e-06;
  Matches   76;  Conservative   77;  Mismatches  140;  Indels   55;  Gaps   10;

Qy      68 QEEENVLDAEFLKNELDSVKAQLSQKDREKRDSQAIIDTLRDTLEERNATVESLQNALNK 127
           | ::   | | | :  | |   : : :: ||  :  :: :|  |||    ::: ::|  :
Db    1517 qnkqlradmedlmsskddvgknvhelekskraleqqveemrtqleeledelqatedaklr 1576

Qy     128 AEMLCSTLKKQMKFLEQRQDE---------TKQAREEAHRLKCKMK-------TMEQIEL 171
            |:    :| | :   | :||          || ||    |: : |       : :::|:
Db    1577 levnmqamkaqferdlqtrdeqneekkrllikqvreleaelederkqralavaskkkmei 1636

Qy     172 LLQSQRSEVE--EMIRDMGVGQSAVEQLAVYCVSLKKEYENLKEARKATGELADRLKKDL 229
            |:   :::|     ||        ::||      :|     |:||| :  |:  : |:
Db    1637 dlkdleaqieaankard-----evikqlrklqaqmkdyqreleearasrdeifaqske-- 1689

Qy     230 VSSRSKLKTLNTELDQAKLELRSAQKDLQSADQEITSLRKKSDDPPGNLEPASATNETVS 289
             |  |||:|  |: | : || |:::  : |:||   |   :|:   :    ||  :
Db    1690 --sekklksleaeilqlqeelasserarrhaeqerdel---adeitnsasgksalldekr 1744

Qy     290 RL----------VFESPAPVEMMNPRLHQPPFGDEIDLNTTFDVNTPPTQTSGSQHCLPK 339
           ||          : |  : :|::| |  :           ||  |:|   : :  :    |
Db    1745 rleariaqleeeleeeqsnmellndrfrk----------ttlqvdtlnaelaaersaaqk 1794

Qy     340 ----KLCLERARSPMQNVLKKVHKVSKPESQLSLGG-QRCVGELDEEL 382
                :  |||    ::  |:::    | : : ::   :  :|:|:|:|
Db    1795 sdnarqqlerqnkelkaklqelegavkskfkatisaleakigqleeql 1842


RESULT 12
ABB71396
ID   ABB71396 standard; Protein; 482 AA.
XX
AC   ABB71396;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 40980.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL15499.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 40980; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
```

```
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   482 AA;


  Query Match             7.4%;  Score 178;  DB 22;  Length 482;
  Best Local Similarity   20.0%;  Pred. No. 1.3e-06;
  Matches   74;  Conservative   56;  Mismatches   96;  Indels  144;  Gaps   14;

Qy       4 LSLCTICSDFFDHSRD-VAAIHCGHTFHLQCLIQWFETAPSRTCPQCR------------ 50
Db     172 lptcpvclermdesvdgvltilcnhafhasclmkwgds----tcpvcrhvqtpglvedsv 227

Qy      51 ----------------------------------------IQVGKKTI--------INK 61
Db     228 cmecegtdslwiclicghvgcgryqgghaaahfratnhtfamqlgtssvwdyagdnfvhr 287

Qy      62 LFFDLAQEEENVLDAEFLKNELDSVKAQLSQKDREKRDSQAIIDTLRDTLEERNATVESL 121
Db     288 lf-------qnksdgklv--------asqtekd------------------ereekidsm 314

Qy     122 QNALNKAEMLCSTLKKQMKFLEQRQDETKQA----REEAHRLKCKMKTMEQIELLLQSQR 177
Db     315 q--meftylltsqldtqrkyyeermerleqewqnhkatandaktevselqqlqqnmqkek 372

Qy     178 SEVEEMIRDMGVGQSAVEQLAVYCVSLKKEYENLKEARKATGELADRLKKDLVSSRSKLK 237
Db     373 vnle-------------rklaqhtaklkdvqkqlneer----elskalqsnqsswhgkyk 415

Qy     238 TLNTELDQAKLELRSAQKDLQSADQEITSLRKKSDD---------PPGNLEPASATNETV 288
Db     416 lleqqynefk----------qthdaevtelkeqlrdimffldnqqklanteiagg---tv 462

Qy     289 SRLVFESPAP 298
Db     463 tgiaekepdp 472


RESULT  13
ABG01723
ID   ABG01723 standard; Protein; 1851 AA.
XX
AC   ABG01723;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #1714.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS65910.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 20; SEQ ID No 32082; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1851 AA;


  Query Match             7.4%;  Score 177;  DB 22;  Length 1851;
  Best Local Similarity   21.5%;  Pred. No. 1.1e-05;
  Matches  109;  Conservative   83;  Mismatches  188;  Indels  128;  Gaps   19;

Qy      12 DFFDHSRDVAAIHCGHTFHLQCLIQWFETAPSRTCPQCRIQVGKKTIINKLFFDLAQEEE 71
Db    1049 dstdlsdqiaelqaqhrgaqdaagqerggapgppwpenkaekqkr--------dlgeele 1100

Qy      72 NVLDAEFLKNELDSVKAQLSQKDREKRDSQAIIDTLRDTLEERNATVES-LQNALNKAEM 130
Db    1101 -alkte-ledtldstaaq--qelrskreqe--vnilkktleeeaktheaqiqemrqkhsq 1154

Qy     131 LCSTLKKQMKFLEQRQDETKQAREEAHRLKCKMKTMEQIELLLQSQRSEVEEMIRDMGV- 189
Db    1155 aveelaeql-------eqtkrvkanlekakqtlen-ergelanevkrkkveaqlqelqvk 1206

Qy     190 ---GQSAVEQLAVYCVSLKKEYEN--------------------LKEARKATGELADRL 225
Db    1207 fnegervrteladkvtklqveldnvpgllsqsdskssskltkdfsalesqlqdtqellqee 1266

Qy     226 KKDLVSSRSKLKTLNTELDQAKLELRSAQKDLQSADQEITSLRKKSDDPPGNLEPASATN 285
Db    1267 nrqklslstklkqleeeeeeakhnl---ekqiatlhaqvadmkkkmedsvgcletaeevk 1323

Qy     286 -------ETVSRLVFESPA---PVEMMNPRLHQPPFGDEIDLN------------TTFD 322
Db    1324 rklqkdleglsqrheekvaaydklektktrlqqelddllvdldhqrqsacnlekkqkkfd 1383

Qy     323 VNTPPTQTSGSQH--------------------------CLPKKLCLERA----RSPMQ 351
Db    1384 qllaeektisakyaeerdraeaeareketkalslaraleeameqkaelerlnkqfrteme 1443

Qy     352 NVL-------KKVHKVSKPESQLSLGGQRC---VGELDEELAGAFPLFIRNAV------- 394
Db    1444 dlmsskddvgksvhelekskraleqqveemktqleeledelqatedaklrlevnlqamka 1503

Qy     395 ------LGQKQPNRTTAESRSSTDVVRI 416
Db    1504 qferdlqgrdepergeeeaagqtaikql 1531


RESULT  14
AAM78854
```

```
ID   AAM78854 standard; Protein; 1960 AA.
XX
AC   AAM78854;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human protein SEQ ID NO 1516.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorder; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US04098.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   27-APR-2000; 2000US-0560875.
PR   20-JUN-2000; 2000US-0598075.
PR   19-JUL-2000; 2000US-0620325.
PR   01-SEP-2000; 2000US-0654936.
PR   15-SEP-2000; 2000US-0663561.
PR   20-OCT-2000; 2000US-0693325.
PR   30-NOV-2000; 2000US-0728422.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT,  Asundi V,  Zhou P,  Xu C,  Cao Y,  Ma Y;
PI   Zhao QA,  Wang D,  Wang J,  Zhang J,  Ren F,  Chen R,  Wang ZW;
PI   Xue AJ,  Yang Y,  Wejhrman T,  Goodrich R;
XX
DR   WPI; 2001-476283/51.
DR   N-PSDB; AAK51987.
XX
PT   Nucleic acids encoding polypeptides with cytokine-like activities,
PT   useful in diagnosis and gene therapy -
XX
PS   Claim 20; Page 3813-3817; 6221pp; English.
XX
CC   The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC   encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC   (AAM80020) are omitted as the relevant pages from the sequence listing
CC   were missing at the time of publication.
XX
SQ   Sequence   1960 AA;


  Query Match             7.4%;  Score 177;  DB 22;  Length 1960;
  Best Local Similarity   19.0%;  Pred. No. 1.2e-05;
  Matches  107;  Conservative  114;  Mismatches  181;  Indels  160;  Gaps   21;

Qy      48 QCRIQVGKKTIINKLFFDLAQEEENV-----------LDAEF--LKNELDSVKAQLSQK 93
           : ::|: ||    :|   ||: ||              |:::   |: :|:| :|  ::
Db    1073 elkmqlakke--eelqaalarveeeaaqknmalkkirelesqiselqedleserasrnka 1130

Qy      94 DREKRDSQAIIDTLRDTLEERNATVESLQNALNKAEMLCSTLKK---------QMKFLEQ 144
           :::|||    :: |:  ||:   :  : |   :| |   : |||         : :  |
Db    1131 ekqkrdlgeelealkteledtldstaaqqelrskreqevnilkktleeeaktheaqiqem 1190

Qy     145 RQ------DETKQAREEAHRLKCKM----KTME--------QIELLLQS------QRSEV 180
           ||      :|  :  |:  |:|  :    :|:|        ::::|||       :| :|
Db    1191 rqkhsqaveelaeqleqtkrvkanlekakqtlenergelanevkvllqgkgdsehkrkkv 1250

Qy     181 EEMIRDMGV----GQSAVEQLAVYCVSLKKEYENLKEARKATGELADRLKKDLVSSRSKL 236
           |  :::: |    |:    :||     |: | :|:      :   : :| ||  :  |:|
Db    1251 eaqlqelqvkfnegervrteladkvtklqveldnvtgllsqsdsksskltkdfsalesql 1310

Qy     237 K--------------TLNTELDQAKLELRSAQKDLQSADQ--------------EITSLR 268
           :              :|:|:| | : |  | :: |:  ::              ::  ::
Db    1311 qdtqellqeenrqklslstklkqvedeknsfreqleeeeeakhnlekqiatlhaqvadmk 1370

Qy     269 KKSDDPPGNLEPASATN-------ETVSRLVFESPA---PVEMMNPRLHQPPFGDEIDLN 318
           || :|  | || |           | :|:   |  |    :|    || |      :||:
Db    1371 kkmedsvgcletaeevkrklqkdleglsqrheekvaaydklektktrlqqeldd1lvdld 1430

Qy     319 -------------TTFDVNTPPTQTSGSQH--------------------------CLP 338
                        ||      :|  :::                          :
Db    1431 hqrqsacnlekkqkkfdqllaeektisakyaeerdraeaeareketkalslaraleeame 1490

Qy     339 KKLCLERA----RSPMQNVL-------KKVHKVSKPESQLSLGGQRC---VGELDEELAG 384
           :|  |||     |: |::::       | ||:: | :  |    :     : ||::||
Db    1491 qkaelerlnkqfrtemedlmsskddvgksvhelekskraleqqveemktqleeledelqa 1550

Qy     385 AFPLFIRNAVLGQKQPNRTTAESRSSTDVVRIGFDGLGGRTKFIQPRDTTIIRPV-PVKS 443
             :|  |      |    :::   |        | || :  : :   ::| | :::
Db    1551 tedaklrlev------nlqamkaqferd--------lqgrdeqseekkkqlvrqvremea 1596

Qy     444 KAKSKQKVRIKTVSSASQPKLD 465
           : : ::| |   |::  : ::|
Db    1597 elederkqrsmavaarkklemd 1618


RESULT  15
ABG01716
ID   ABG01716 standard; Protein; 2143 AA.
XX
AC   ABG01716;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #1707.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS65903.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 32075; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
```

```
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   2143 AA;


  Query Match             7.4%;  Score 177;  DB 22;  Length 2143;
  Best Local Similarity   19.0%;  Pred. No. 1.4e-05;
  Matches  107;  Conservative  114;  Mismatches  181;  Indels  160;  Gaps   21;

Qy      48 QCRIQVGKKTIINKLFFDLAQEEENV-----------LDAEF--LKNELDSVKAQLSQK 93
           : ::|: ||    :|   ||: ||             |:::   |: :|:| :|  ::
Db    1155 elkmqlakke--eelqaalarveeeaaqknmalkkirelesqiselqedleserasrnka 1212

Qy      94 DREKRDSQAIIDTLRDTLEERNATVESLQNALNKAEMLCSTLKK---------QMKFLEQ 144
           :::|||    :: |:  ||:   :  : |   :| |   : |||         : :  |
Db    1213 ekqkrdlgeelealkteledtldstaaqqelrskreqevnilkktleeeaktheaqiqem 1272

Qy     145 RQ------DETKQAREEAHRLKCKM----KTME--------QIELLLQS------QRSEV 180
           ||      :|  :  |:  |:|  :    :|:|        ::::|||       :| :|
Db    1273 rqkhsqaveelaeqleqtkrvkanlekakqtlenergelanevkvllqgkgdsehkrkkv 1332

Qy     181 EEMIRDMGV----GQSAVEQLAVYCVSLKKEYENLKEARKATGELADRLKKDLVSSRSKL 236
           |  :::: |    |:    :||     |: | :|:      :   : :| ||  :  |:|
Db    1333 eaqlqelqvkfnegervrteladkvtklqveldnvtgllsqsdskssktkdfsalesql 1392

Qy     237 K--------------TLNTELDQAKLELRSAQKDLQSADQ-------------EITSLR 268
           :             :|:|:| | : |  | :: |:  ::             ::  ::
Db    1393 qdtqellqeenrqklslstklkqvedeknsfreqleeeeeakhnlekqiatlhaqvadmk 1452

Qy     269 KKSDDPPGNLEPASATN-------ETVSRLVFESPA---PVEMMNPRLHQPPFGDEIDLN 318
           || :|  | || |           | :|:   |  |    :|    || |      :||:
Db    1453 kkmedsvgcletaeevkrklqkdleglsqrheekvaaydklektktrlqqelddllvdld 1512

Qy     319 ------------TTFDVNTPPTQTSGSQH-------------------------CLP 338
                       ||      :|  :::                          :
Db    1513 hqrqsacnlekkqkkfdqllaeektisakyaeerdraeaeareketkalslaraleeame 1572

Qy     339 KKLCLERA----RSPMQNVL-------KKVHKVSKPESQLSLGGQRC---VGELDEELAG 384
           :|  |||     |: |::::       | ||:: | :  |    :     : ||::||
Db    1573 qkaelerlnkqfrtemedlmsskddvgksvhelekskraleqqveemktqleeledelqa 1632

Qy     385 AFPLFIRNAVLGQKQPNRTTAESRSSTDVVRIGFDGLGGRTKFIQPRDTTIIRPV-PVKS 443
                :|  |       |    :::   |        | || :  : :   ::| |  :::
Db    1633 tedaklrlev------nlqamkaqferd--------lqgrdeqseekkkqlvrqvremea 1678

Qy     444 KAKSKQKVRIKTVSSASQPKLD 465
           : : ::| |   |::  : ::|
Db    1679 elederkqrsmavaarkklemd 1700


RESULT  16
ABB62094
ID   ABB62094 standard; Protein; 1975 AA.
XX
AC   ABB62094;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 13074.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL06197.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 13074; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1975 AA;


  Query Match             7.4%;  Score 176.5;  DB 22;  Length 1975;
  Best Local Similarity   22.4%;  Pred. No. 1.3e-05;
  Matches   87;  Conservative   68;  Mismatches  139;  Indels   95;  Gaps   15;

Qy      48 QCRIQVGK-KTIINKLFFDLAQEEENVLDAEFLKNELDSVKAQLSQKDREKRDS-QAIID 105
           :|  ::|| : ::  |     |||:: |: :  | :: :::| | :| : : |  | : :
Db    1462 rcdaklgklqamlrnl-----qeeksnletd-rkmkisaiqa-leeklkhrndecqmlre 1514

Qy     106 TLRDTLEERNATVE---SLQNALNKAEMLCSTLKKQMKFLEQRQDETKQAREEAHRLKCK 162
             |  |  :  || |     :  | |:   || |  : :   | |:|  :    | :|: :
Db    1515 rlaqtemqlaatseengqneerleksrqqcskldnekr---qlqeelakvegrasklelq 1571

Qy     163 MKTME----QIELLLQSQRSEVEEMIRDMGVGQSAVEQLAVYCVSLKKEYENLKEARKAT 218
             ||    :::: || :   : :|   :     |: ||   | :||   : |||
Db    1572 rvamegdltrlqmalqekdcsirqmaerlenqnraltqledrctalkstvdqlke----- 1626

Qy     219 GELADRLKKDLVSS---RSKLKTLNTELDQA-------KLELRSAQKDLQSADQEITSLR 268
               ||:|  ||    | ::|||  || :        : :|:  || ||:|: |   |
Db    1627 -----rlqksavsetqlrgeiktlqkelseqghcsqanedklklvqkslqtaenekrilt 1681

Qy     269 KKSDDPPGNLEPASATNETVSRLVFESPAPVEMMNPRLHQPPFGDEIDLNTTFDVNTPPT 328
           :: |    ||     : :
Db    1682 erldsaqtnlnelrrsqq----------------------------------------a 1700

Qy     329 QTSGSQHCLPKKLCLERARSPMQNVLK--KVHKVSKPESQLSLGGQRCVGELDEELAGAF 386
```

```
          |  |:|    :   ||  || ::: |:  | :: |  :  |: |     ||  |||:
Db   1701 qldgnqrlqeqvtdlevqrsalesqlriakwnqesggdkdltngngggnge--eelsrql 1758

Qy    387 PLFIRNAVLGQKQPNRTTAESRSSTDVVR 415
                  |   |  :| ||    ::
Db   1759 ----------kssqrekselrsklqtlq 1776


RESULT  17
AAE02246
ID   AAE02246 standard; Protein; 1017 AA.
XX
AC   AAE02246;
XX
DT   31-JUL-2001  (first entry)
XX
DE   Domestic mite Bt11 allergen polymorphic variant.
XX
KW   Mite; immunogenic protein; Bt allergen; therapy; atopic dermatitis;
KW   immediate hypersensitivity; systemic anaphylaxis; allergic rhinitis;
KW   asthma; antiallergic; antiinflammatory; immunosuppressive.
XX
OS   Blomia tropicalis.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 41
FT                   /note= "Encoded by TAG"
FT   Misc-difference 42
FT                   /note= "Encoded by TAG"
FT   Misc-difference 56
FT                   /note= "Encoded by TGA"
FT   Misc-difference 71
FT                   /note= "Encoded by TAA"
FT   Misc-difference 76
FT                   /note= "Encoded by TAG"
FT   Misc-difference 80
FT                   /note= "Encoded by TGA"
FT   Misc-difference 86
FT                   /note= "Encoded by TAA"
FT   Misc-difference 965
FT                   /note= "Encoded by TAA"
FT   Misc-difference 998
FT                   /note= "Encoded by TAA"
XX
PN   WO200130817-A1.
XX
PD   03-MAY-2001.
XX
PF   10-OCT-2000; 2000WO-AU01227.
XX
PR   26-OCT-1999;   99SG-0005313.
PR   18-JUL-2000; 2000AU-0008842.
PR   18-JUL-2000; 2000AU-0008844.
PR   18-JUL-2000; 2000AU-0008845.
XX
PA   (UYSI-) UNIV SINGAPORE NAT.
XX
PI   Chua KY,  Cheong N,  Lee BW;
XX
DR   WPI; 2001-308609/32.
DR   N-PSDB; AAD06245.
XX
PT   Novel immunogenic protein derived from house mite, Blomia tropicalis
PT   useful for treating and diagnosing conditions involving induction of
PT   immuneresponse to mite, such as allergic asthma, atopic dermatitis,
PT   rhinitis  -
XX
PS   Claim 6; Fig 7; 230pp; English.
XX
CC   The present invention relates to immunogenic proteins, referred as Bt
CC   allergen, is derived from domestic mite, Blomia tropicalis. The specific
CC   Bt allergens of the invention includes Bt11, Bt10, Bt5 and BtA2. The
CC   immunogenic protein is useful for preventing, reducing or ameliorating
CC   Blomia tropicalis hypersensitivity condition such as atopic dermatitis,
CC   immediate hypersensitivity, systemic anaphylaxis, allergic rhinitis or
CC   asthma and for modulating an immune response directed to Bt allergen in
CC   a subject. The Bt allergens are also useful for detecting antibody
CC   directed to all or a part of Bt allergen in a biological sample from a
CC   subject. Antibodies to Bt allergens are also used as therapeutic or
CC   diagnostic agents, to screen Bt immunoassays and as antagonists to
CC   inhibit Bt activity under circumstances where temporary hypersensitivity
CC   inhibition is required. The present sequence is a protein encoded
CC   by Bt11 polymorphic variant.
XX
SQ   Sequence   1017 AA;


  Query Match             7.3%;  Score 175.5;  DB 22;  Length 1017;
  Best Local Similarity   22.0%;  Pred. No. 6.2e-06;
  Matches   90;  Conservative   71;  Mismatches  152;  Indels   97;  Gaps   15;

Qy     76 AEFLKNELDSVKAQL----------------------SQKDREKRDSQAIIDTL----RD 109
          |  |: ||:|:| ||                      | | : : : || :| :    |
Db    346 ahtleveleslkvqleeesearlelerqltkangdaaswkskyeaelqahvdeveelrrk 405

Qy    110 TLEERNATVESLQNALNKAEMLCSTLKKQMKFLEQRQD----ETKQAREEAHRLKCKMKT 165
            :: :   | |:  |||    || |:||   |:   :    : ::|   |  |: ::
Db    406 maqkiseyeeqleallnk----csalekqkarlqseaevlimdlekatahaqalekrvsq 461

Qy    166 MEQIELLLQSQRSEVEEMIRDMGVGQSAVEQLAVYCVSLKK---EYENLKEARKATGELA 222
          :|:| | |:|:  ||  ::      :   : | |    |:|   ||| |:: ::|
Db    462 lekinldlkskleevsmll------eqtqkdlrvkiadlqklqheyeklrdqkealaren 515

Qy    223 DRLKKDLVSSRSKLKTLNTELDQAKLE---LRSAQKDLQSADQEITSLRKKSDDPPGNLE 279
           :|  ||  ::|:|   :  : : ::|   | : :::| :| :|  :|||: :     |
Db    516 kkladdlaeaksqlndahrriheqeieikrlenereelaaaykeaetlrkqeeaknqrlt 575

Qy    280 PASA------------TNETVSRLVFESPAPVEMMNPRLHQPPFGDEIDLNTTFDVNTPP 327
            |             | :  |  :    :| :| || :     |  | |
Db    576 aelaqtrhdyekrlaqkeeeiealrkqyqieieqlnmrlaea----eaklkt----evar 627

Qy    328 TQTSGSQHCLPKKLCLE---RARSPMQNVLKK-------------VHK-VSKPESQLSL 369
            :          :| |:   :|   :|  :||             ||: : :   || :
Db    628 lkkkyqaqitelelsldaankanidlqktikkqalqitelqahydevhrqlqqavdqlgv 687

Qy    370 GGQRC---VGELDEELAGAFPLFIRNAVLGQKQPNRTTAESRSSTDVVRI 416
             :||     ||:|             | |    :  ||      |||:
Db    688 tqrrcqaltaelee----------mrvnleqalrakraaeqmheeavvrv 727


RESULT  18
AAB96721
ID   AAB96721 standard; Protein; 1177 AA.
XX
AC   AAB96721;
XX
DT   29-OCT-2001  (first entry)
XX
DE   Putative P. abyssi ATPase involved in DNA repair #3.
XX
KW   Hyperthermophilic archaeon; hyperthermophilic protein.
XX
OS   Pyrococcus abyssi.
XX
PN   FR2792651-A1.
XX
PD   27-OCT-2000.
XX
PF   21-APR-1999;   99FR-0005034.
XX
PR   21-APR-1999;   99FR-0005034.
XX
PA   (CNRS ) CNRS CENT NAT RECH SCI.
PA   (IFRE-) IFREMER INST FR RECH EXPL MER.
XX
PI   Forterre P,  Thierry JC,  Prieur D,  Dietrich J,  Lecompte O;
```

```
PI   Querellou J,  Weissenbach J,  Saurin W,  Heilig R;
XX
DR   WPI; 2001-126236/14.
XX
PT   New nucleotide sequences isolated from Pyrococcus abyssi encode
PT   proteins useful in industry -
XX
PS   Claim 7; Pages 1483-1487; 1657pp; French.
XX
CC   The present invention relates to the genomic sequence of Pyrococcus
CC   abyssi (see AAF86431 and AAH41223-7) and P. abyssi proteins. P. abyssi is
CC   a hyperthermophilic archaeon, which is isolated from deep-sea
CC   hydrothermal vents. The present sequence is one such P. abyssi protein.
CC   The proteins of the present invention have various potential industrial
CC   uses, since the proteins are stable at very high temperatures, some up to
CC   110 degrees centigrade.
CC   Note: This patent is in the same patent family as WO200065062, which
CC   contains additional sequences as shown in AAB99132-AAB99143,
CC   AAH75903-AAH75920 and AAG66436.
XX
SQ   Sequence   1177 AA;


  Query Match             7.3%;  Score 175;  DB 22;  Length 1177;
  Best Local Similarity   21.0%;  Pred. No. 8.5e-06;
  Matches   88;  Conservative   86;  Mismatches  155;  Indels   90;  Gaps   17;

Qy      61 KLFFDLAQEEENVLDAEFLKNELDSVKAQLSQKDREKRDSQAIIDTLRDTLEERNATVES 120
           |   :| | |||:   : |  |   || || : ::|: |:   :| |:: ||   | ||
Db     178 kalqelkqaeenlarvdllire---vkkqldklekerndalryld-lkerle--rarvel 231

Qy     121 LQNALNKAEMLCSTLKKQMKFLEQRQDETKQAREE----AHRLKCKMKTMEQIELLLQSQ 176
           :   : | |   | :|   : :|: ::| |:  |:    |  :  | : ::::| |:: :
Db     232 ilgeikkve---seikgnderiekieeeikeieekleeiakeivrkerelkeveelieke 288

Qy     177 RS-----------EVEEMI----RDMGVGQSAVEQLAVYCVSLKKEYEN-LKEARKATGE 220
             |           ||   |    |:: | :  :::  :  :  | | :  | |  |: |
Db     289 sseealkitreigevnskinlakrnievakkeldeaqirlikakdelkkvlseiekskga 348

Qy     221 LADRLK---------KDLVSSRSKLKTLNTELDQAKLELRSAQKDLQSADQEITSLRKKS 271
           :|   |         |:|   |:||     |:|:       |:::  :  :|: : ||
Db     349 iarwgkrkeallnkikeleeernklvvklgeidrt---favareefdnvvkelenarksl 405

Qy     272 DDPPGNLEPASATNETVSRLVFESPAPVEMMNPRLHQPPFGDEIDLNTTFDVNTPPTQTS 331
           :   :::   | | :|       : : ::  :|  |   :|::
Db     406 yeneadikrleaekerls-------sritilkakl--pgireevek-------------- 442

Qy     332 GSQHCLPKKLCLERARSPMQNVLKKVHKVSKPESQLSLGGQRCVGELDE---ELAGAFPL 388
             | :|  ||  :: : ||  |:  :|:   ::    ::   || :   ||
Db     443 -----lrek--leekkaelsnvenkissisqrrrkveeelekktselqkvsseleslere 495

Qy     389 FIRNAVLGQKQPNRTTAESRSSTDVVRIGFDGLGGRTKFIQPRDTTIIRPVPVKSKAKS 447
           |:     : : ||   | :      | |  |: |          |::  : || :  |
Db     496 likaeaqsevrvnraveelk------rsgisgiyg----------tllelirvkdemys 538


RESULT  19
AAE02245
ID   AAE02245 standard; Protein; 875 AA.
XX
AC   AAE02245;
XX
DT   31-JUL-2001  (first entry)
XX
DE   Domestic mite Bt11 allergen.
XX
KW   Mite; immunogenic protein; Bt allergen; therapy; atopic dermatitis;
KW   immediate hypersensitivity; systemic anaphylaxis; allergic rhinitis;
KW   asthma; antiallergic; antiinflammatory; immunosuppressive.
XX
OS   Blomia tropicalis.
XX
PN   WO200130817-A1.
XX
PD   03-MAY-2001.
XX
PF   10-OCT-2000; 2000WO-AU01227.
XX
PR   26-OCT-1999;   99SG-0005313.
PR   18-JUL-2000; 2000AU-0008842.
PR   18-JUL-2000; 2000AU-0008844.
PR   18-JUL-2000; 2000AU-0008845.
XX
PA   (UYSI-) UNIV SINGAPORE NAT.
XX
PI   Chua KY,  Cheong N,  Lee BW;
XX
DR   WPI; 2001-308609/32.
DR   N-PSDB; AAD06237.
XX
PT   Novel immunogenic protein derived from house mite, Blomia tropicalis
PT   useful for treating and diagnosing conditions involving induction of
PT   immuneresponse to mite, such as allergic asthma, atopic dermatitis,
PT   rhinitis -
XX
PS   Disclosure; Page 162-166; 230pp; English.
XX
CC   The present invention relates to immunogenic proteins, referred to as Bt
CC   allergens, derived from domestic mite (Blomia tropicalis). The specific
CC   Bt allergens of the invention includes Bt11, Bt10, Bt5 and BtA2. The
CC   immunogenic protein is useful for preventing, reducing or ameliorating
CC   Blomia tropicalis hypersensitivity condition such as atopic dermatitis,
CC   immediate hypersensitivity, systemic anaphylaxis, allergic rhinitis or
CC   asthma and for modulating an immune response directed to Bt allergen in
CC   a subject. The Bt allergens are also useful for detecting antibody
CC   directed to all or a part of Bt allergen in a biological sample from a
CC   subject. Antibodies to Bt allergens are also used as therapeutic or
CC   diagnostic agents, to screen Bt immunoassays and as antagonists to
CC   inhibit Bt activity under circumstances where temporary hypersensitivity
CC   inhibition is required. The present sequence is Bt11 allergen.
XX
SQ   Sequence   875 AA;


  Query Match             7.3%;  Score 174.5;  DB 22;  Length 875;
  Best Local Similarity   21.7%;  Pred. No. 6.1e-06;
  Matches   89;  Conservative   71;  Mismatches  153;  Indels   97;  Gaps   15;

Qy      76 AEFLKNELDSVKAQL----------------------SQKDREKRDSQAIIDTLRDTLEE 113
           |  |: ||:|:| ||                      | | : : : || :| : :   :
Db     257 ahtleveleslkvqleeesearlelerqltkangdaaswkskyeaelqahvdeveelrrk 316

Qy     114 RNATV----ESLQNALNKAEMLCSTLKKQMKFLEQRQD----ETKQAREEAHRLKCKMKT 165
                :    | |:  |||    || |:||   |:   :    : ::|   |  |: ::
Db     317 maqkiseygeqleallnk----csalekqkarlqsevevlimdlekatahaqalekrvsq 372

Qy     166 MEQIELLLQSQRSEVEEMIRDMGVGQSAVEQLAVYCVSLKK---EYENLKEARKATGELA 222
           :|:| | |:|:  ||  ::      :   : | |    |:|   ||| |:: ::|
Db     373 lekinldlkskleevsmll------eqtqkdlrvkiadlqklqheyeklrdqkealaren 426

Qy     223 DRLKKDLVSSRSKLKTLNTELDQAKLE---LRSAQKDLQSADQEITSLRKKSDDPPGNLE 279
           :|  ||  ::|:|   :  : : ::|   | : :::| :| :|  :|||: :     |
Db     427 kkladdlaeaksqlndahrriheqeieikrlenereelaaaykeaetlrkqeeaknqrlt 486

Qy     280 PASA-----------TNETVSRLVFESPAPVEMMNPRLHQPPFGDEIDLNTTFDVNTPP 327
             |              | :  |  :    :| :| || :     |  | |
Db     487 aelaqtrhdyekrlaqkeeeiealrkqyqieieqlnmrlaea----eaklkt----evar 538

Qy     328 TQTSGSQHCLPKKLCLE---RARSPMQNVLKK-------------VHK-VSKPESQLSL 369
           :          :| |:   :|  :|  :||               ||: : :   || :
Db     539 lkkkyqaqitelelsldaankanidlqktikkqalqitglqahydevhrqlqqavdqlgv 598

Qy     370 GGQRC---VGELDEELAGAFPLFIRNAVLGQKQPNRTTAESRSSTDVVRI 416
             :||     ||:|             | |    :  ||      |||:
Db     599 tqrrcqaltaelee----------mrvnleqalrakraaeqmheeavvrv 638
```

```
RESULT  20
AAE02242
ID   AAE02242 standard; Protein; 878 AA.
XX
AC   AAE02242;
XX
DT   31-JUL-2001  (first entry)
XX
DE   Domestic mite Bt11 allergen #7.
XX
KW   Mite; immunogenic protein; Bt allergen; therapy; atopic dermatitis;
KW   immediate hypersensitivity; systemic anaphylaxis; allergic rhinitis;
KW   asthma; antiallergic; antiinflammatory; immunosuppressive.
XX
OS   Blomia tropicalis.
XX
PN   WO200130817-A1.
XX
PD   03-MAY-2001.
XX
PF   10-OCT-2000; 2000WO-AU01227.
XX
PR   26-OCT-1999;   99SG-0005313.
PR   18-JUL-2000; 2000AU-0008842.
PR   18-JUL-2000; 2000AU-0008844.
PR   18-JUL-2000; 2000AU-0008845.
XX
PA   (UYSI-) UNIV SINGAPORE NAT.
XX
PI   Chua KY,  Cheong N,  Lee BW;
XX
DR   WPI; 2001-308609/32.
DR   N-PSDB; AAD06236.
XX
PT   Novel immunogenic protein derived from house mite, Blomia tropicalis
PT   useful for treating and diagnosing conditions involving induction of
PT   immuneresponse to mite, such as allergic asthma, atopic dermatitis,
PT   rhinitis  -
XX
PS   Claim 4; Fig 3; 230pp; English.
XX
CC   The present invention relates to immunogenic proteins, referred as Bt
CC   allergen, is derived from domestic mite, Blomia tropicalis. The specific
CC   Bt allergens of the invention includes Bt11, Bt10, Bt5 and BtA2. The
CC   immunogenic protein is useful for preventing, reducing or ameliorating
CC   Blomia tropicalis hypersensitivity condition such as atopic dermatitis,
CC   immediate hypersensitivity, systemic anaphylaxis, allergic rhinitis or
CC   asthma and for modulating an immune response directed to Bt allergen in
CC   a subject. The Bt allergens are also useful for detecting antibody
CC   directed to all or a part of Bt allergen in a biological sample from a
CC   subject. Antibodies to Bt allergens are also used as therapeutic or
CC   diagnostic agents, to screen Bt immunoassays and as antagonists to
CC   inhibit Bt activity under circumstances where temporary hypersensitivity
CC   inhibition is required. The present sequence is Bt11 allergen.
XX
SQ   Sequence   878 AA;


  Query Match             7.3%;  Score 174.5;  DB 22;  Length 878;
  Best Local Similarity   21.7%;  Pred. No. 6.1e-06;
  Matches   89;  Conservative   71;  Mismatches  153;  Indels   97;  Gaps   15;

Qy      76 AEFLKNELDSVKAQL----------------------SQKDREKRDSQAIIDTLRDTLEE 113
           |  |: ||:|:| ||                      | | : : : || :| : :   :
Db     260 ahtleveleslkvqleeesearlelerqltkangdaaswkskyeaelqahvdeveelrrk 319

Qy     114 RNATV----ESLQNALNKAEMLCSTLKKQMKFLEQRQD----ETKQAREEAHRLKCKMKT 165
               :    | |:  |||    || |:||   |:   :    : ::|   |  |: ::
Db     320 maqkiseygeqleallnk----csalekqkarlqsevevlimdlekatahaqalekrvsq 375

Qy     166 MEQIELLLQSQRSEVEEMIRDMGVGQSAVEQLAVYCVSLKK---EYENLKEARKATGELA 222
           :|:| | |:|:  ||  ::      :   : | |    |:|   ||| |:: ::|
Db     376 lekinldlkskleevsmll------eqtqkdlrvkiadlqklqheyeklrdqkealaren 429

Qy     223 DRLKKDLVSSRSKLKTLNTELDQAKLE---LRSAQKDLQSADQEITSLRKKSDDPPGNLE 279
             :|  ||  ::|:|   :  : : ::|   | : :::| :| :|  :|||: :     |
Db     430 kkladdlaeaksqlndahrriheqeieikrlenereelaaaykeaetlrkqeeaknqrlt 489

Qy     280 PASA----------TNETVSRLVFESPAPVEMMNPRLHQPPFGDEIDLNTTFDVNTPP 327
             |             | :  |  :    :| :| || :     |  | |
Db     490 aelaqtrhdyekrlaqkeeeieialrkqyqieieqlnmrlaea----eaklkt----evar 541

Qy     328 TQTSGSQHCLPKKLCLE---RARSPMQNVLKK-------------VHK-VSKPESQLSL 369
           :          :| |:   :|   :|  :||              ||: : :   || :
Db     542 lkkkyqaqitelelsldaankanidlqktikkqalqitglqahydevhrqlqqavdqlgv 601

Qy     370 GGQRC---VGELDEELAGAFPLFIRNAVLGQKQPNRTTAESRSSTDVVRI 416
             :||     ||:|             | |    :  ||      |||:
Db     602 tqrrcqaltaelee----------mrvnleqalrakraaeqmheeavvrv 641


RESULT  21
AAB69070
ID   AAB69070 standard; Protein; 1374 AA.
XX
AC   AAB69070;
XX
DT   19-APR-2001  (first entry)
XX
DE   Human male enhanced antigen-2 (MEA-2) protein sequence SEQ ID NO:2.
XX
KW   Human; male enhanced antigen-2; MEA-2; identification; spermatogenesis;
KW   spermatogenesis disease; chromosome marker; pancreatic cancer.
XX
OS   Homo sapiens.
XX
PN   JP2000316580-A.
XX
PD   21-NOV-2000.
XX
PF   30-APR-1999;   99JP-0125196.
XX
PR   30-APR-1999;   99JP-0125196.
XX
PA   (ITOH-) ITO HAM KK.
XX
DR   WPI; 2001-128256/14.
DR   N-PSDB; AAF32508.
XX
PT   A new protein, human male-enhanced antigen-2, useful for detecting
PT   spermatogenesis diseases  -
XX
PS   Claim 1; Page 12-15; 21pp; Japanese.
XX
CC   The present sequence represents the human male enhanced antigen-2
CC   (MEA-2). The present invention also described an antibody specific for
CC   the MEA-2 protein. The antibody can be used for the identification of a
CC   gene causing diseases related to spermatogenesis. The MEA-2 nucleotide
CC   sequence is useful as a chromosome marker, and in the detection of
CC   pancreatic cancer.
XX
SQ   Sequence   1374 AA;


  Query Match             7.3%;  Score 174;  DB 22;  Length 1374;
  Best Local Similarity   23.0%;  Pred. No. 1.3e-05;
  Matches   96;  Conservative   68;  Mismatches  153;  Indels  100;  Gaps   15;

Qy      66 LAQEEENVLDAEFL-----KNELDSVKAQLSQKDREKRDSQAIIDTLRDTLEERNATVES 120
           | | |: :|:|        | |||    :| :   |  ::   :: ||: |  ::  ||
Db     754 llqnekiileaalqaaksgkeeldrgarrleegteetset---leklreelaiksgqveh 810

Qy     121 LQNALNKAEMLCSTLKKQM-----KFLEQR-------------------------QDE 148
           ||          : |||||     :||:|:                         |
Db     811 lqqe-------taalkkqmqkikeqflqqkvmveayrrdatskdqliselkatrkrldse 863
```

```
Qy    149 TKQAREEAHRLKCKMKT-----------------MEQIELLLQSQRSEVEEMIRDMGVG 190
          |: |:|  ::  : :|                  |  :|  ||| : | :||   :
Db    864 lkelrqelmqvhgekrtaeaelsrlhrevaqvrqhmadleghlqsaqkerdemethlqsl 923

Qy    191 QSAVEQLAVYC---VSLKKEYENL-KEARKATGELADRLKK---DLVSSRSKLKTLNTEL 243
          |   ||:        :|||: | | :|||||  |   ::::   || |:: ::|| :
Db    924 qfdkeqmvavteanealkkqieelqqearkaiteqkqkmrrlgsdltsaqkemktkhkay 983

Qy    244 DQA----KLELRSAQKDLQSADQEITSLRKKSDDPPGNLEPASATNETVSRLVFESPAPV 299
          : |       |: |    ::|| |:  || :     |: : : | :| :  |  |  | |
Db    984 enavgilsrrlqealaakeaadaelgqlraqg----gssdsslalheriqaleaelqa-v 1038

Qy    300 EMMNPRLHQPPFGDEIDLNTTFDVNTPPTQTSGSQHCLPKKLCLERARSPMQNVLKKVHK 359
              |       | :|     : :   : |       | | ||      :   ||: :
Db   1039 shsktll-------ekelqevialtsqeleesre-----kvleledelqesrgfrkkikr 1086

Qy    360 VSKPESQLSLGGQRCVGELDEELAGAFPLFIRNAVLGQKQPNRTTAESRSSTDVVRI 416
          : :   :|:|       ||: |      |   || | :      || ::   |:|::
Db   1087 leesnkklal-------elehekgkltglgqsnaalrehnsiletalakreadlvql 1136


RESULT  22
ABB59948
ID   ABB59948 standard; Protein; 1489 AA.
XX
AC   ABB59948;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 6636.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL04051.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 6636; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1489 AA;
```

```
  Query Match             7.3%;  Score 174;  DB 22;  Length 1489;
  Best Local Similarity   21.4%;  Pred. No. 1.4e-05;
  Matches   98;  Conservative   60;  Mismatches  163;  Indels  138;  Gaps   14;

Qy     56 KTIINKLFFDLAQEEENVLDAEFLKNELDSVKAQLSQKDREKRDSQAIIDTLRDTLEERN 115
          |::  |      |||||||    |     ::  :|  || | |  |  :  |    |  |
Db    298 ksvtekyeavrkqeeenvl---llaqtkqaihtelelkdtevrklqeklkqlesqreshn 354

Qy    116 -----------ATVESLQNALNKAEMLCSTLKKQMKFLEQR------------------- 145
                      || : :   |   | | :|||:     ||:
Db    355 nevkeqfkklqatkqevdaklmatehllntlkesyaikeqqvvtleaqleairveneqkv 414

Qy    146 -------QDETKQAREEAHRLKCKMKTMEQIELLLQSQRSEVEEMIRDMGVGQSAVEQLA 198
                 :|   || : : :||     ::  |  | |:   :| :  :    :  :: |
Db    415 kdlqkqnedrntqasdsseqlkklqaavqdaesqllskdqlleslrseqaakeqqlkhlk 474

Qy    199 VYCVSLKKEYEN----LKEARKA----TGELADRLKK----------------DLVSS-- 232
               ||:| ||    |:| :|:    | |  |: ||                 :|: |
Db    475 eqlgklkqenenyldklrenkkssdsqtneaqdqqkklqaakdeaesklltateellhslr 534

Qy    233 -------------RSKLKTLNTELD--QAKLELRSAQKDLQSAD--QEITSLRKKSDDPP 275
                        |||||: | |    ||   : |:: || |  |:|  ||   |:
Db    535 ndykaqeekvalledklktlskendvnveklhhineqreaqstdsqqkinelraakdeae 594

Qy    276 GNL-----------------EPASATNETVSRLVFESPAPVEMMNPRLHQPPFGDEIDL 317
            |                  | |::  :::: |  ||   :: :  |||     : :
Db    595 akllstehslnalqaalsakeeqaasleqslnalktesehslqdl--rlhndqlleivqr 652

Qy    318 NTTFD---------VNTPPTQTSGSQHCLPKKLCLERARSPMQ---------------- 351
          :   |              |:    | |  :  ||  |  :
Db    653 hqqndweaqlarareelaaiqsqrelhaleleklemeresvaalnsekasqeeqhrlkl 712

Qy    352 -------NVLKKVHKVSKPESQLSLGGQRCVGELDEELA 383
                 :|:  |  |: |:  :| ||  :  | ::||
Db    713 eqlqreiqilqdqhansesetvaalkgq--lealsqdla 749


RESULT  23
AAM79264
ID   AAM79264 standard; Protein; 1203 AA.
XX
AC   AAM79264;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human protein SEQ ID NO 1926.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorder; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US04098.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   27-APR-2000; 2000US-0560875.
PR   20-JUN-2000; 2000US-0598075.
PR   19-JUL-2000; 2000US-0620325.
PR   01-SEP-2000; 2000US-0654936.
PR   15-SEP-2000; 2000US-0663561.
PR   20-OCT-2000; 2000US-0693325.
PR   30-NOV-2000; 2000US-0728422.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT,  Asundi V,  Zhou P,  Xu C,  Cao Y,  Ma Y;
```

```
PI   Zhao QA,  Wang D,  Wang J,  Zhang J,  Ren F,  Chen R,  Wang ZW;
PI   Xue AJ,  Yang Y,  Wejhrman T,  Goodrich R;
XX
DR   WPI; 2001-476283/51.
DR   N-PSDB; AAK52397.
XX
PT   Nucleic acids encoding polypeptides with cytokine-like activities,
PT   useful in diagnosis and gene therapy -
XX
PS   Claim 20; Page 4327-4329; 6221pp; English.
XX
CC   The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC   encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC   (AAM80020) are omitted as the relevant pages from the sequence listing
CC   were missing at the time of publication.
XX
SQ   Sequence   1203 AA;


  Query Match             7.3%;  Score 173.5;  DB 22;  Length 1203;
  Best Local Similarity   21.0%;  Pred. No. 1.2e-05;
  Matches   96;  Conservative   80;  Mismatches  181;  Indels  101;  Gaps   15;

Qy      17 SRDVAAIHCGHTFHLQCLIQWFETAPSRTCPQCRIQVGKKTIINKLFFDLAQEEENVLDA 76
           |::||  |       |  :   |    |   :  :|: |     :   : | : : | :|
Db     650 sqevagrhrdrelekqlavlrveadrgreleeqnlqlqktlqqlrqdceeaskakmvaea 709

Qy      77 EF---------------LKNELDSVKAQLSQKDREKRDSQAIID-------TLRDTLEE 113
           |                 : | |  : ::   ::: :::: ::|        ||| |:
Db     710 eatvlgqrraavettlretqeendefrrrilgleqqlketrglvdggeavearlrdklqr 769

Qy     114 RNATVESLQNALNKAEMLCSTLKKQMKFLEQRQDETKQ--AR--EEAHRLKCKMKTMEQI 169
           |  : |: ||| ::    :|    : || | :| ::  ||  :|   |   ::   :
Db     770 leaekqqleealnasqeeegslaaakralearleeaqrglarlgqeqqtlnraleeegkq 829

Qy     170 ELLLQSQRSEVEEMIRDMG-----------VGQSAVEQLAVYCVSLKKEYENLKEARKA 217
             :|:  ::|:|| | :             :|: : : |      |: : |: ||  ||
Db     830 revlrrgkaeleeqkrlldrtvdrlnkelekigedskqalq----qlqaqledyke--ka 883

Qy     218 TGELAD-------------------RLKKDLVSSRSKLKTLNTELDQAKLELRSAQKDL 257
           |:||                    ||: ::   |  |:    | | |:|:     : |
Db     884 rrevadaqrqakdwaseaektsgglsrlqdeiqrlrqalqasqaerdtarldkellaqrl 943

Qy     258 QSADQEITSLRKKSDDPPGNLEPASATNETVSRL---VFESPAPVEMMNPRLHQPPFGDE 314
           |  :|| : ::  ||    |:      | ||||   : |    ||::  |:::   |:
Db     944 qgleqeaenkkrsqddrarqlk---gleekvsrleteldeekntvelltdrvnrg--rdq 998

Qy     315 ID-LNTTFDVNTPPTQTSGSQHCLPKKLCLERARSPMQNVLKKVHKVSKPESQLSLGGQR 373
           :| | |         : |  |     |: |||    ::  |       || : ||
Db     999 vdqlrtel-----mqersarqdlecdkislerqnkdlktrlassegfqkpsasls----- 1048

Qy     374 CVGELDEELAGAFPLFIRNAVLGQKQPNRTTAESRSST 411
             :|: :               |    |  || |  |
Db    1049 ---qlesq-------------nqllqerlqaeerekt 1069


RESULT  24
AAM40292
ID   AAM40292 standard; Protein; 864 AA.
XX
AC   AAM40292;
XX
DT   22-OCT-2001  (first entry)

XX
DE   Human polypeptide SEQ ID NO 3437.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   N-PSDB; AAI59448.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Example 5; SEQ ID NO 3437; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence   864 AA;


  Query Match             7.2%;  Score 172.5;  DB 22;  Length 864;
  Best Local Similarity   20.0%;  Pred. No. 8.9e-06;
  Matches   92;  Conservative   75;  Mismatches  179;  Indels  115;  Gaps   15;

Qy      82 ELDSVKAQLSQKDREKRDSQAIIDTLRDTLEERNATVESLQNALNKAEMLCSTLKKQMKF 141
           ||| :  :::|  |||   :  |    : : :: : |: ||| |::       |: | :
Db     385 elddisqeiaqlqrekysleqdirekeeairqktsevqelqndldretsslqeleaqkqd 444

Qy     142 LEQRQDETKQAREEAH------RLKCKMKT--MEQIELLLQSQRSEVEEMIRDMGVGQSA 193
           : | ||  | : :        | ||: :|  :  ::  :||| |:::    |:   :|
Db     445 aqdrldemdqqkaklrdmlsdvrqkcqdetqmisslktqiqsqesdlksqeddlnrakse 504
```

```
Qy    194 VEQLAVYCVSLKKEYENLKEARKATGELADRLKKDLVSSRSKLKTLNTELDQAKLELRSA 253
          : :       |::|   |::: :|     : : | | |:: ::    ::| |     : |
Db    505 lnr-------lqqeetqleqsiqagrvqletiikslkstqdeinqarsklsqlhesrqea 557

Qy    254 QKDLQSADQ--------EITSLRKKS--------------DDPPGNLEPASATNETVSRL 291
           : |:  ||         :| |   |               |||  |             |
Db    558 hrsleqydqvldgahgasltdlanlsegvslaergsfgamddpfkn-----------kal 606

Qy    292 VFES------PAPVEMMNPRLHQP-------------------------PFG------- 312
          :| :      | | :  :|   |                        |||
Db    607 lfsnntqelhpdpfqtedpfksdpfkgadpfkgdpfqndpfaeqqttstdpfggdpfkes 666

Qy    313 DEIDLNTTFDV------NTPPTQTSGSQH-CLPKKLCLERARSPMQNVLKKVHKVSKPES 365
          |    : | |       | | |    :::  || ||    :  |  :            |
Db    667 dpfrgsatddffkkqtkndpftsdpftknpslpskldtfessdpfss-----------s 714

Qy    366 QLSLGGQRCVGELDEELAGAFPLFIRNAVLGQKQPNRTTAESRSSTDVVRIGFDGLGGRT 425
           :|  |    | ||   :|:|     |:  |    :: :    |      :|  |
Db    715 svsskgsdpfgtldpfgsgsf-----nsaegfadfsqmskpppsgpftsslggagfsd-d 768

Qy    426 KFIQPRDTTII---RPVPVKSKAKSKQKVRIKTVSSASQPK 463
           |   :||  :   :| | : |  | :   :  : ||  |:
Db    769 pfkskqdtpalppkkpapprpkppsgkstpvsqlgsadfpe 809


RESULT  25
AAR72826
ID   AAR72826 standard; Protein; 2482 AA.
XX
AC   AAR72826;
XX
DT   27-FEB-1996  (first entry)
XX
DE   Human mitosin.
XX
KW   Cell cycle; M phase; mitosin; retinoblastome; mitosis; cell growth;
KW   inhibition.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          1480..1659
FT                   /label= internal_repeat
FT   Region          1660..1839
FT                   /label= internal_repeat
XX
PN   WO9511309-A2.
XX
PD   27-APR-1995.
XX
PF   24-OCT-1994;   94WO-US12162.
XX
PR   22-OCT-1993;   93US-0141239.
XX
PA   (TEXA ) UNIV TEXAS SYSTEM.
XX
PI   Lee W,  Zhu X;
XX
DR   WPI; 1995-170229/22.
DR   N-PSDB; AAQ86851.
XX
PT   Purified mammalian protein mitosin and agents that bind it and
PT   inhibit its action  - used to promote cell growth or to inhibit cell
PT   division and/or proliferation
XX
PS   Claim 4; Fig 8B; 61pp; English.
XX
CC   AAR72829 is human mitosin. Mitosin is involved in the regulation of
CC   the mammalian mitotic cell cycle. Mitosin as with E2F-1 (see AAR72824)
CC   interacts with the retinoblastoma protein (the retinoblastoma tumour
CC   suppressor gene product). Mitosin is first synthesised at the G1/S
CC   boundary, it is then phosphorylated from S through M phase, and during
CC   mitosis, is closely associated with the centromeres/kinetochores at the
CC   mitotic spindle poles. Mitosin is necessary for a eukaryotic cell to
CC   enter the M phase of the mitotic cell cycle and its degradation is
CC   necessary for a cell to advance on to the next stage. Mitosin is thus
CC   useful for controlling cell growth as overexpression of mitosin prevents
CC   a cell from exiting the M phase.
CC   An anti-mitosin antibody, antibody fragment or a phosphorylated mitosin
CC   mutein ( or nucleic acid encoding it) can also be used to inhibit cell
CC   division which is particularly useful for the study of the cell cycle.
CC   A further use is to control hyperproliferative cells, and so control
CC   diseases such as psoriasis and breast cancer. It can also be used to
CC   block gametogenesis of an immature gamete.
XX
SQ   Sequence   2482 AA;


  Query Match             7.2%;  Score 172.5;  DB 16;  Length 2482;
  Best Local Similarity   24.3%;  Pred. No. 4.1e-05;
  Matches   74;  Conservative   53;  Mismatches  109;  Indels   69;  Gaps   10;

Qy     56 KTIINKLFFDLAQEEEN----VLDAEFLKNELDSVKAQL-----------------SQK 93
          |  :  |  :|   |||    :|||| | |::::| |:                  |:|
Db   1521 kdkvenlerelqmseenqelvildaenskaevetlktqieemarslkvfeldlvtlrsek 1580

Qy     94 D------REKRDSQAIIDTLRDTL--------------EERNATVESLQNALNKAEMLC 132
          :      :||:   : :| |  :                ||    || ||| | :
Db   1581 enltkqiqekqgqlseldkllssfkslleekeqaeiqikeesktavemlqnqlkelneav 1640

Qy    133 STL---KKQMKFLEQRQD----ETKQAREEAHRLKCKMKTMEQIELLLQSQRSE------ 179
          : |   :: ||  ||  |     |  | |    :|: :::  |: :| :  |  |
Db   1641 aalcgdqeimkateqsldppieeehqlrnsieklrarleadekkqlcvlqqlkesehhad 1700

Qy    180 -----VEEMIRDMGVGQSAVEQLAVYCVSLKKEYENLKEARKATGELADRLKKDLVSSRS 234
               || : |:: : ::  |  |:   : | | | ||   :   :    |: |:|: ||
Db   1701 llkgrvenlereleiartnqehaaleaenskgevetlkakiegmtqslrgleldvvtirs 1760

Qy    235 KLKTLNTELDQ-----AKLEL--RSAQKDLQSADQEITSLRKKSDDPPGNLE-PASATNE 286
          : : |  || :      ::||:   | :  ||  :||   :::||      |:      ||
Db   1761 ekenltnelqkeqeriseleiinssfenilqekeqekvqmkeksstamemlqtqlkelne 1820

Qy    287 TVSRL 291
           |: |
Db   1821 rvaal 1825


RESULT  26
AAW23996
ID   AAW23996 standard; Protein; 2482 AA.
XX
AC   AAW23996;
XX
DT   28-MAY-1998  (first entry)
XX
DE   Human mitosin amino acid sequence.
XX
KW   Mitosin; phosphoprotein; mitotic cell cycle; antibody; analogue;
KW   inhibition; M phase; Antagonist; hyperproliferative cell; cancer;
KW   leukaemia; lymphoma; chromosome segregation.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          258..280
FT                   /note= "leucine heptad repaet"
FT   Domain          340..362
FT   Domain          564..593
FT   Domain          1387..1443
FT   Domain          1885..1962
FT   Domain          2146..2188
FT   Domain          2165..2187
FT                   /note= "leucine heptad repeat"
FT   Misc-difference 2188
FT   Misc-difference 2300
FT                   /label= "Bipartite targeting motif"
```

```
FT                   /note= "Optionally C or G"
FT   Misc-difference 2189
FT   Misc-difference 2301
FT   Misc-difference 2303
FT                   /label= "Bipartite targeting motif"
FT                   /note= "optinally A or T"
XX
PN   US5710022-A.
XX
PD   20-JAN-1998.
XX
PF   24-OCT-1994;   94US-0328254.
XX
PR   24-OCT-1994;   94US-0328254.
PR   22-OCT-1993;   93US-0141239.
XX
PA   (TEXA ) UNIV TEXAS SYSTEM.
XX
PI   Lee W,  Zhu X;
XX
DR   WPI; 1998-109817/10.
DR   N-PSDB; AAV09076.
XX
PT   New isolated mitosin protein and gene - useful for, e.g. developing
PT   products for therapy and diagnosis of hyper-proliferative disorders
PT   such as cancers or psoriasis
XX
PS   Claim 1; Column 40-52; 43pp; English.
XX
CC   This is the amino acid sequence for mitosin, a phosphoprotein
CC   necessary for the cell to ente mitosis.  The protein's degradation is
CC   also necessary for the cell to advance into the next stages of mitosis.
CC   The mitosin protein, can be used to control the growth of cells.  An
CC   anti-mitosin antibody, a mutant or a non-functional analogue of mitosin
CC   can inhibit the mitotic cell cycle by preventing the cells from entering
CC   the M phase, and over expression of mitosin or its functional
CC   equivalent, would inhibit the cycle by preventing cells from leaving the
CC   M phase. Antagonists to this protein can be used to control
CC   hyperproliferative cells in, (e.g. thyroid hyperplasia, Grave's disease,
CC   psoriasis, benign prostatic hypertrophy, Li-Fraumeni syndrome, breast
CC   cancer, sarcomas and other neoplasms, bladder cancer, colon cancer,
CC   lung cancer and various leukaemias and lymphomas).  Reintroduction or
CC   supplementation of lost mitosin function by introduction of the protein
CC   or nucleic acid encoding the protein into a cell can restore defective
CC   chromosome segregation, which is a marker of progressing malignancy.
CC   Malignant proliferation of cells can then be halted.  The protein
CC   can also be used for the detection and diagnosis of hyperproliferative
CC   cells.
XX
SQ   Sequence   2482 AA;


  Query Match             7.2%;  Score 172.5;  DB 19;  Length 2482;
  Best Local Similarity   24.3%;  Pred. No. 4.1e-05;
  Matches   74;  Conservative   53;  Mismatches  109;  Indels   69;  Gaps   10;

Qy      56 KTIINKLFFDLAQEEEN----VLDAEFLKNELDSVKAQL------------------SQK 93
           |  :  |  :|   |||    :|||| | |::::| |:                   |:|
Db    1521 kdkvenlerelqmseenqelvildaenskaevetlktqieemarslkvfeldlvtlrsek 1580

Qy      94 D------REKRDSQAIIDTLRDTL--------------EERNATVESLQNALNKAEMLC 132
           :      :||:   : :| |  :                ||    || ||| | :
Db    1581 enltkqiqekqgqlseldkllssfkslleekeqaeiqikeesktavemlqnqlkelneav 1640

Qy     133 STL---KKQMKFLEQRQD----ETKQAREEAHRLKCKMKTMEQIELLLQSQRSE------ 179
           : |   :: ||  ||  |    |  | |    :|: ::::  |: :| :  |  |
Db    1641 aalcgdqeimkateqsldppieeehqlrnsieklrarleadekkqlcvlqqlkesehhad 1700

Qy     180 -----VEEMIRDMGVGQSAVEQLAVYCVSLKKEYENLKEARKATGELADRLKKDLVSSRS 234
                || : |:: : ::  |  |:   : | | | ||   :   :    |: |:|: ||
Db    1701 llkgrvenlereleiartnqehaaleaenskgevetlkakiegmtqslrgleldvvtirs 1760

Qy     235 KLKTLNTELDQ-----AKLEL--RSAQKDLQSADQEITSLRKKSDDPPGNLE-PASATNE 286
```

```
           : : |  || :     ::||:   | :  ||  :||   :::||      |:      ||
Db    1761 ekenltnelqkeqeriseleiinssfenilqekeqekvqmkeksstamemlqtqlkelne 1820

Qy     287 TVSRL 291
           |: |
Db    1821 rvaal 1825


RESULT  27
AAR99795
ID   AAR99795 standard; Protein; 3248 AA.
XX
AC   AAR99795;
XX
DT   08-OCT-1996  (first entry)
XX
DE   Kinetochore protein CENP-F.
XX
KW   Kinetochore protein; CENP-F; cell cycle; cancer; diagnosis;
KW   autoimmune antibody.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          1..200
FT                   /label= Extended_coiled_structure
FT   Domain          280..1350
FT                   /label= Extended_coiled_structure
FT   Domain          1380..1610
FT                   /label= Globular_domain
FT                   /note= "globular domain consists of 2 direct
FT                         repeats of 95 amino acids"
FT   Domain          1620..1750
FT                   /label= Extended_coiled_structure
FT   Domain          1850..2990
FT                   /label= Extended_coiled_structure
FT   Domain          3048..3248
FT                   /label= C-terminal_domain
FT                   /note= "the C-terminal domain is predicted to
FT                         form a proline-rich (10.6%) highly
FT                         basic (pI 10) globular domain"
XX
PN   WO9617867-A1.
XX
PD   13-JUN-1996.
XX
PF   08-DEC-1995;   95WO-US16216.
XX
PR   09-DEC-1994;   94US-0353700.
XX
PA   (FOXC-) FOX CHASE CANCER CENT.
PA   (UYTE-) UNIV TECHNOLOGIES INT INC.
XX
PI   Rattner JB,  Yen TJ;
XX
DR   WPI; 1996-287116/29.
DR   N-PSDB; AAT34578.
XX
PT   DNA encoding kinetochore protein - used as a marker for the G2 and M
PT   phases of a cell cycle, partic. for detection of malignant diseases
XX
PS   Claim 12; Page 41-54; 72pp; English.
XX
CC   A 372 kDa human kinetochore protein, CENP-F (AAR99795), is detected
CC   by immunofluorescence microscopy only during the G2 and M phases
CC   of a cell cycle.  It is the product of a cDNA clone (AAT34578)
CC   isolated from a breast carcinoma cDNA library.  Recombinant CENP-F
CC   can be produced by expression in prokaryotic or eukaryotic host
CC   cells.  CENP-F can be used to detect autoimmune antibodies to
CC   the protein, which may provide an early diagnosis for the onset
CC   of various malignant diseases.  Use of CENP-F as a cell cycle
CC   marker allows the specific detection of G2 and M phase cells.
XX
```

```
SQ   Sequence   3248 AA;

  Query Match             7.2%;  Score 172.5;  DB 17;  Length 3248;
  Best Local Similarity   24.3%;  Pred. No. 6e-05;
  Matches   74;  Conservative   53;  Mismatches  109;  Indels   69;  Gaps   10;

Qy      56 KTIINKLFFDLAQEEEN----VLDAEFLKNELDSVKAQL----------------SQK 93
           |  :  |  :|   |||    :|||| | |::::| |:                  |:|
Db    2249 kdkvenlerelqmseenqelvildaenskaevetlktqieemarslkifeldlvtlrsek 2308

Qy      94 D------REKRDSQAIIDTLRDTL--------------EERNATVESLQNALNKAEMLC 132
           :      :||:   : :| |  :                ||    || ||| | :
Db    2309 enltkqiqekqgqlseldkllssfkslleekeqaeiqikeesktavemlqnqlkelneav 2368

Qy     133 STL---KKQMKFLEQRQD----ETKQAREEAHRLKCKMKTMEQIELLLQSQRSE------ 179
           : |   :: ||  ||  |    |  | |    :|: :::  |: :| :  |  |
Db    2369 aalcgdqeimkateqsldppieeehqlrnsieklrarleadekkqlcvlqqlkesehhad 2428

Qy     180 -----VEEMIRDMGVGQSAVEQLAVYCVSLKKEYENLKEARKATGELADRLKKDLVSSRS 234
                || : |:: : ::  |  |:   : | | | ||   :   :    |: |:|: ||
Db    2429 llkgrvenlereleiartnqehaaleaenskgevetlkakiegmtqslrgleldvvtirs 2488

Qy     235 KLKTLNTELDQ-----AKLEL--RSAQKDLQSADQEITSLRKKSDDPPGNLE-PASATNE 286
           : : |  || :     ::||:   | :  ||  :||   :::||      |:      ||
Db    2489 ekenltnelqkeqeriseleiinssfenilqekeqekvqmkeksstamemlqtqlkelne 2548

Qy     287 TVSRL 291
           |: |
Db    2549 rvaal 2553


RESULT  28
AAB21231
ID   AAB21231 standard; Protein; 717 AA.
XX
AC   AAB21231;
XX
DT   09-MAR-2001  (first entry)
XX
DE   Tomato LeMPF1.
XX
KW   Tomato; MAR binding filament-like protein 1; MFP1; LeMFP1;
KW   matrix attachment region; MAR; NtMFP1-2; anchor protein.
XX
OS   Lycopersicon esculentum.
XX
PN   WO200061615-A2.
XX
PD   19-OCT-2000.
XX
PF   12-APR-2000; 2000WO-US09723.
XX
PR   12-APR-1999;   99US-0128900.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Harder PA,  Meier I;
XX
DR   WPI; 2000-679464/66.
DR   N-PSDB; AAA95812.
XX
PT   Nucleic acid fragments from tobacco, corn, soybean and rice, encoding
PT   proteins that are homologs to the MAR binding filament-like protein 1
PT   (MFP1), useful for development of novel phenotypes -
XX
PS   Claim 5; Page 54-56; 62pp; English.
XX
CC   The present sequence is LeMFP1, the matrix attachment region (MAR)
CC   binding filament-like protein 1 (MFP1) from tomato. MFP1
CC   has features of a novel anchor protein that most likely connects
CC   chromatin via MAR DNA with the nuclear envelope and nuclear filament
CC   proteins. MFP1 nucleic acids and proteins may be used to better
CC   understand the mechanisms underlying this process so that the attachment
CC   of transgenes to the nuclear matrix may be used routinely to improve gene
CC   expression. They may be used to study MFP1 expression, leading to the
CC   creation of novel developmental phenotypes that may be beneficial for
CC   crop growth and development. In addition, if the reduction in expression
CC   of one of the genes leads to a growth or developmental defect in the
CC   plant, this gene can be used as a novel herbicide target.
XX
SQ   Sequence   717 AA;

  Query Match             7.2%;  Score 172;  DB 21;  Length 717;
  Best Local Similarity   20.2%;  Pred. No. 7.5e-06;
  Matches   76;  Conservative   69;  Mismatches  138;  Indels   94;  Gaps   10;

Qy      62 LFFDLAQEEEN---VLDAEF-----LKNELDSVKAQLSQKDREKRDSQAIIDTLRDTLEE 113
           |  || :|:||   ::|||      || |::  :  |     |  |    :: |||   :
Db     395 liadltrekenlrrmvdaeldnvnklkqeievtqeslensrsevsditvqleqlrdlcsk 454

Qy     114 RNATVESLQNALNKAEMLCSTLKKQMKFLEQRQDETKQAREEAHRLKCKMKTMEQIELLL 173
             | |  ||            |::    |::  |||| : |                 ||
Db     455 leaevsklq----------meleetraslqrnidetkhsse----------------ll 487

Qy     174 QSQRSEVEEMIRDMGVGQSAVEQLAVYCVSLKKEYENLKEARKATGELADRLKKDLVSSR 233
             :: :  :|:::                  :|   : :   |  |  | |: :||
Db     488 aaeltttkellk----------------ktneemhtmsdelvavsenrdslqtelvdvy 530

Qy     234 SKLKTLNTELDQAKLELRSAQKDLQSADQEIT---SLRKKSDDPPGNLEPASATNETVSR 290
           | :    || | |  : : :::|:  : :||    |||  :|    || |: : : ::|
Db     531 kkaehtanelkqeksivatleeelkflesqitrekelrksled---elekatesldeinr 587

Qy     291 LVFESPAPVEMMNPRLHQPPFGDEIDLNTTFDVNTPPTQTSGSQHCLPKKLCLERARSPM 350
           |                     :|::| |: : :    :    |    :|   : |: :
Db     588 nvl----------------alaeelelatsrnsslederevlrqsvseqkqisqeaqenl 631

Qy     351 QNVLKKVHKVSKPESQLSLGGQRCVGELDEELAGAFPLFIR----------------NAV 394
           ::    | |: |    |    ::   :|::|:| |    :|                |
Db     632 edahslvmklgkeresl----ekrakkledemaaakgeilrlrsqinsvkapvedeekvv 687

Qy     395 LGQKQPNRTTAESRSST 411
           |:|:  : |  :: :|
Db     688 agekekvkatvtakktt 704


RESULT  29
ABB58673
ID   ABB58673 standard; Protein; 1456 AA.
XX
AC   ABB58673;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 2811.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
```

```
DR   N-PSDB; ABL02776.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 2811; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1456 AA;
```

```
  Query Match             7.2%;  Score 171.5;  DB 22;  Length 1456;
  Best Local Similarity   21.1%;  Pred. No. 2.3e-05;
  Matches  111;  Conservative   88;  Mismatches  221;  Indels  107;  Gaps   18;

Qy      15 DHSRDVAAIH---CGHTFHLQCLIQWFETAPSRTCPQCRIQVGKKTIINKLFFDLAQEEE 71
           |: | :| :    |    |   |    |   :|   : |: : |||             :
Db      55 dyqrhiavlkeslcakeehynmlqtdveemrarleeknrl-iekkt---------qgtlq 104

Qy      72 NVLDAEFLKNELDSVKAQLSQKDREKRDSQAIIDTLRDTLEERNATVE------------ 119
             |  :   | :||  :|  :  |||:    |  |: | | |:|::  |:
Db     105 tvqernrltseltelkdhmdikdrkisvlqrkienledllkekdnqvdmararlsamqah 164

Qy     120 --SLQNALNKAEMLCSTLKKQMKFLEQRQDETKQAREEAHRLKCKMKTMEQIELLLQSQR 177
             | : ||   |      :|||  | ::|  :  ::|   |    : :   :: |::
Db     165 hsssegaltsleeaigdkekqmaqlrdqrdraehekqeerdl--herevadykiklraae 222

Qy     178 SEVEEMIRDMGVGQSAVEQLAVYCVSLKKEYENLK-EARKATGELA-------------- 222
           ||||::   :    :  |:| :   : : |    | |  ||| |:
Db     223 seveklqtrleravtererleikleasqselgkskaelekatcemgrssadwestkqria 282

Qy     223 ------DRLKKDLVSSRSKL----KTLNTELDQA-------------KLELRSAQKDLQ 258
                :||| ||  |::       | : |||:|              : |||  | | :
Db     283 rlelenerlkhdlersqttfgrttmttsqeldraqeradkasaelrrtqaelrvtqsdae 342

Qy     259 SADQEITSLRKKSDDPPG-------NLEPASATNETVSRLVFESPAPVEMMNPRLH---Q 308
             | :|  :|::| :   |         || |    |:: : : :: : |     |:|
Db     343 rareeaaalqekleksqgevyrlkaklenaqgeqeslrqelekaqsgvs----rihadrd 398

Qy     309 PPFGDEIDLNTTFDVNTPPTQTSGSQHCLPKKL--CLERARSPMQNVLKKVHKVSKPESQ 366
             | :   :    :        |  ||   :||   |::|:: : ::  |: |      :
Db     399 rafsevekikeemertqatlgksqlqh---eklqnsldkaqnevdhlqdkldkactenrr 455

Qy     367 LSLGGQRCV-------GELDEELAGAFPLFIRNAVLGQKQPNRTTAESRSSTDVVRIGFD 419
           | |  ::          :||: |  |           : |  | |      || :|   :
Db     456 lvlekekltydydnlqsqldkalgqa----------armqkeretls--ldtdrirekle 503

Qy     420 GLGGRTKFIQ-PRDTTIIRPVPVKSKAKSKQKVRIKTVSSASQPKLD 465
             :   ||  ||        :| :::| | : :|       : |
Db     504 ktqvqlgriqkerdqfsdeletlkersesaqtllmkaardreamqtd 550
```

```
RESULT  30
AAY86194
ID   AAY86194 standard; Protein; 866 AA.
XX
AC   AAY86194;
XX
DT   11-APR-2000  (first entry)
XX
DE   Nuclear transport protein clone hfb2230 protein sequence.
XX
KW   Nuclear transport protein; drug delivery system; visual detection; human;
KW   nuclear transport indicator.
XX
OS   Homo sapiens.
XX
PN   WO9964455-A1.
XX
PD   16-DEC-1999.
XX
PF   04-JUN-1999;   99WO-JP03015.
XX
PR   05-JUN-1998;   98JP-0174065.
PR   14-APR-1999;   99JP-0107262.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ueki N,  Yano K;
XX
DR   WPI; 2000-105872/09.
DR   N-PSDB; AAZ96752.
XX
PT   Peptides with nuclear transport activity, applicable as carrier for
PT   transport desirable substances, e.g. drugs into nucleus, and as nuclear
PT   transport indicators after modification for visual detection -
XX
PS   Claim 1; Page 62-67; 309pp; Japanese.
XX
CC   This sequence represents the amino acid sequence of a nuclear transport
CC   protein of the invention. The invention relates to peptides with nuclear
CC   transport activity, and also includes similar peptides with the same
CC   transport activity, but with some amino acids deleted, substituted and/or
CC   added. The peptides can be used as carriers for transferring desirable
CC   substances particularly drugs, including nucleic acids and proteins, into
CC   the nucleus in a drug delivery system. The peptides can also be used as
CC   nuclear transport indicators after modification for visual detection.
CC   Drugs can be delivered efficiently with the peptides to a specific target
CC   site. This delivery allows the maintenance of drug concentration,
CC   therapeutic efficacy and a reduction of side-effects of such drugs to be
CC   achieved.
XX
SQ   Sequence   866 AA;
```

```
  Query Match             7.1%;  Score 170.5;  DB 21;  Length 866;
  Best Local Similarity   20.0%;  Pred. No. 1.3e-05;
  Matches   92;  Conservative   75;  Mismatches  179;  Indels  115;  Gaps   15;

Qy      82 ELDSVKAQLSQKDREKRDSQAIIDTLRDTLEERNATVESLQNALNKAEMLCSTLKKQMKF 141
           ||| :  :::|  |||   :  |    : : :: : |: ||| |::       |: | :
Db     387 elddisqeiaqlqrekysleqdirekeeairqktsevqelqndldretsslqeleaqkqd 446

Qy     142 LEQRQDETKQAREEAH------RLKCKMKT--MEQIELLLQSQRSEVEEMIRDMGVGQSA 193
             : | ||  | : :        | ||: :|  :  ::  :||| |:::    |:   :|
Db     447 aqdrldemdqqkaklrdmlsdvrqkcqdetqmisslktqiqsqesdlksqeddlnrakse 506

Qy     194 VEQLAVYCVSLKKEYENLKEARKATGELADRLKKDLVSSRSKLKTLNTELDQAKLELRSA 253
           : :       |::|   |::: :|     : : | | |:: ::    ::| |     : |
Db     507 lnr-------lqqeetqleqsiqagrvqletiikslkstqdeinqarsklsqlhesrqea 559

Qy     254 QKDLQSADQ--------EITSLRKKS--------------DDPPGNLEPASATNETVSRL 291
             : |:  ||         :| |   |              |||  |             |
Db     560 hrsleqydqvldgahgasltdlanlsegvslaergsfgamddpfkn-----------kal 608

Qy     292 VFES------PAPVEMMNPRLHQP-------------------------PFG------- 312
           :| :       | | :  :|    |                           |||
Db     609 lfsnntqelhpdpfqtedpfksdpfkgadpfkgdpfqndpfaeqqttstdpfggdpfkes 668

Qy     313 DEIDLNTTFDV------NTPPTQTSGSQH-CLPKKLCLERARSPMQNVLKKVHKVSKPES 365
           |    : | |       | | |    :::  || ||    :  |  :           |
Db     669 dpfrgsatddffkkqtkndpftsdpftknpslpskldpfessdpfss-----------s 716

Qy     366 QLSLGGQRCVGELDEELAGAFPLFIRNAVLGQKQPNRTTAESRSSTDVVRIGFDGLGGRT 425
```

```
          :|  |    | ||   :|:|     |:  |    :: :    |      :|  |
Db    717 svsskgsdpfgtldpfgsgsf-----nsaegfadfsqmskpppsgpftsslggagfsd-d 770

Qy    426 KFIQPRDTTII---RPVPVKSKAKSKQKVRIKTVSSASQPK 463
          |   :||  :   :| | : |  | :   :  : ||  |:
Db    771 pfkskqdtpalppkkpapprpkppsgkstpvsqlgsadfpe 811
```

RESULT 31
AAY07031
ID AAY07031 standard; Protein; 1752 AA.
XX
AC AAY07031;
XX
DT 02-JUL-1999 (first entry)
XX
DE Breast cancer associated antigen precursor sequence.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer.
XX
OS Homo sapiens.
XX
PN WO9904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
DR WPI; 1999-132448/11.
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
PS Disclosure; Page 409-413; 787pp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
SQ Sequence 1752 AA;

Query Match 7.1%; Score 170.5; DB 20; Length 1752;
Best Local Similarity 18.0%; Pred. No. 3.6e-05;
Matches 93; Conservative 110; Mismatches 210; Indels 103; Gaps 15;

```
Qy     35 IQWFETAPSRTCPQCRIQVGKKTIINKLFFDLAQEEENVLDAEF---------------- 78
          :|   :: : |  : ::|  : | : :: ::   :|  | | :
Db    395 lqkanssatetinklkvqeqeltrl-ridyervsqertvkdqditrfqnslkelqlqkqk 453

Qy     79 LKNELDSVKAQLSQKDREKRDSQAIIDTLRDTLEERNATVESLQNALNKAEMLC----ST 134
          :: ||: :|   |:   ::: :  :: :| :|:|:   : :|   | :| ::
Db    454 veeelnrlkrtasedsckrkkleeelegmrrslkeqaikitnltqqleqasivkkrsedd 513

Qy    135 LKKQMKFLEQRQDETKQAREEAHRLKCKMKTMEQIELLLQSQRS------------EVE 181
          |::|   |:    | :: :||  ||  ::: : :   ||| | |              :|
Db    514 lrqqrdvldghlrekqrtqeelrrlssevealrr--qllqeqesvkqahlrnehfqkaie 571

Qy    182 EMIRDMGVGQSAVEQLAVYCVSLKKEYENLKE----------ARKATGELADRLKKDLV 230
          :  | :   :  :|:|     :| ||:  |:|           |:   |        ::
Db    572 dksrslneskieierlqsltenltkehlmleeelrnlrleyddlrrgrseadsdknatil 631

Qy    231 SSRSKLKTLNTELDQAKLELRSAQKDLQSADQEITSLRKKSDDPPGNLEPA-----SATN 285
            ||:|:  |    : :  :   |:: ::  |||   :|:: :    :: :
Db    632 elrsqlqisnnrtlelqglindlqrerenlrqeiekfqkqaleasnriqesknqctqvvq 691

Qy    286 ETVSRLV-----FESPAPVEMMNPRLHQPPFGDEIDLNT----------TFDVNTPPTQ 329
          |  | ||       :  | :: :   |::     | :                |:|   ||
Db    692 eresllvkikvleqdkarlqrledelnrakstleaetrvkqrlecekqqiqndlnqwktq 751

Qy    330 TSGSQHCLPKKLCLERARSP-----------MQNVLKKVHK---------VSKPESQLS 368
           |  :  : :|:  || :|             :|  :|:: :           : :|||
Db    752 ysrkeeai-rkieserekserekns lrseierlqaeikrieercrrkledstretqsqle 810

Qy    369 LGGQRCVGELDEELAGAFPLFIRNAVLGQKQPNRTTAESRSSTDVVRIGFDGLGGRTKFI 428
             |   |:|:          :|    |  :  :|  |   : |  :: ||||  :   :
Db    811 tersryqreidk---------lrqrpygshretqt--ecewtvdtsklvfdglrkkvtam 859

Qy    429 QPRDTTIIRPVPVKSKAKSKQKVRIKTVSSASQPKL 464
          |  :  :|    :    | |: |  : |:|  || |
Db    860 qlyecqlidkttldkllkgkksv--eevaseiqpfl 893
```

RESULT 32
ABG21233
ID ABG21233 standard; Protein; 1948 AA.
XX
AC ABG21233;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #21224.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS85420.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX

```
PS   Claim 20; SEQ ID No 51592; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1948 AA;


  Query Match             7.1%;  Score 170.5;  DB 22;  Length 1948;
  Best Local Similarity   20.5%;  Pred. No. 4.2e-05;
  Matches   78;  Conservative   83;  Mismatches  169;  Indels   51;  Gaps   11;

Qy      55 KKTIINKLFFDLAQEEENVLDAEFLKNELDSVKAQLSQKDREKRDSQAIIDTLRDTLEER 114
           :: :||:|    ||:     ::     :||   | :||  | |:     |: |:  |||
Db    1273 qqrlinel---saqkarlhtesgefsrqldekdamvsqlsrgkqaftqqieelkrqleee 1329

Qy     115 NATVESLQNALNKAEMLCSTLKKQMKFLEQRQDETKQAREEAHRLKCKMKTMEQIELL-- 172
             :| :||  |   |  |::| :  :: : | ::   :|:    : :|  : : :
Db    1330 tkakstlahalqsarhdcdllreqyeeeqeakaelqrgmskansevaqwrtkyetdaiqr 1389

Qy     173 ---LQSQRSEVEEMIRDMGVGQSAVEQLAVYCVSLKKEYENLK----------EARKATG 219
             |:  : :: : ::|    :  || :   | ||:|  : |:          |   |
Db    1390 teeleeakkklaqrlqd---aeehveavnskcaslektkqrlqnevedlmidversnaac 1446

Qy     220 ELADRLKK--DLVSSRSKLKTLNTELDQAKLELRSAQKDLQSADQEITSLRKKSDDPPGN 277
             |: ::  | | :  | |      :: : || ::||: :|   |:  ::   ::   :
Db    1447 ialdkkqrnfdkvlaewkqk-----yeetqaeleasqkesrslstelfkvknayeesldh 1501

Qy     278 LEPASATNETVSRLVFESPAPVEMMNPRLHQPPFGDEID-LNTTFDVNTPPTQTSGSQHC 336
           ||     |: : : :    :      :|      |:: :    |      |||
Db    1502 letlkrenknlqqeisdlteqiaeggkhih------elekvkkqldhekselqts----- 1550

Qy     337 LPKKLCLERARSPMQNVLKKVHKVSKPESQLSLGGQRCVGELDEELAGAFPLFIRNAV-L 395
             || | : :::   |: ::    :|:     | : | ||||        || : :
Db    1551 ------leeaeasleheegkilriqlelnqvkseidrkiaekdeel----dqlkrnhlrv 1600

Qy     396 GQKQPNRTTAESRSSTDVVRI 416
           :   :   || ||  | :||
Db    1601 vesmqstldaeirsrndalri 1621


RESULT  33
AAM79838
ID   AAM79838 standard; Protein; 1963 AA.
XX
AC   AAM79838;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human protein SEQ ID NO 3484.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorder; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US04098.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   27-APR-2000; 2000US-0560875.
PR   20-JUN-2000; 2000US-0598075.
PR   19-JUL-2000; 2000US-0620325.
PR   01-SEP-2000; 2000US-0654936.
PR   15-SEP-2000; 2000US-0663561.
PR   20-OCT-2000; 2000US-0693325.
PR   30-NOV-2000; 2000US-0728422.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT,  Asundi V,  Zhou P,  Xu C,  Cao Y,  Ma Y;
PI   Zhao QA,  Wang D,  Wang J,  Zhang J,  Ren F,  Chen R,  Wang ZW;
PI   Xue AJ,  Yang Y,  Wejhrman T,  Goodrich R;
XX
DR   WPI; 2001-476283/51.
DR   N-PSDB; AAK52971.
XX
PT   Nucleic acids encoding polypeptides with cytokine-like activities,
PT   useful in diagnosis and gene therapy -
XX
PS   Claim 20; Page 354-355; 6221pp; English.
XX
CC   The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC   encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC   (AAM80020) are omitted as the relevant pages from the sequence listing
CC   were missing at the time of publication.
XX
SQ   Sequence   1963 AA;


  Query Match             7.1%;  Score 170.5;  DB 22;  Length 1963;
  Best Local Similarity   20.4%;  Pred. No. 4.3e-05;
  Matches  105;  Conservative   99;  Mismatches  167;  Indels  143;  Gaps   22;

Qy      65 DLAQEEENVLDAEFLKNELDSVKAQLSQKDREKRDSQAIIDTLRDTLEERNATVES-LQN 123
           || :| |  |  | |:: |||  ||  |: | ||: :  :: |: ||||   | |: :|
Db    1138 dlgeele-alkte-ledtldstaaq--qelrskreqe--vnilkktleeeaktheaqiqe 1191

Qy     124 ALNKAEMLCSTLKKQMKFLEQRQDETKQAREEAHRLKCKMKTME--------QIELLLQS 175
             |      | :|:       ::||: :    : |   :|:|         ::::|||
Db    1192 mrqkhsqaveelaeql-------eqtkrvkanlekak---qtlenergelanevkvllqg 1241

Qy     176 ------QRSEVEEMIRDMGV----GQSAVEQLAVYCVSLKKEYENLKEARKATGELADRL 225
                :| :||  :::: |    |:    :||     |: | :|:      :   : :|
Db    1242 kgdsehkrkkveaqlqelqvkfnegervrteladkvtklqveldnvtgllsqsdsksskl 1301

Qy     226 KKDLVSSRSKLK--------------TLNTELDQAKLELRSAQKDLQSADQE-------- 263
           ||  :  |:|:               :|:|:| | : |  | :: |:  ::|
Db    1302 tkdfsalesqlqdtqellqeenrqklslstklkqvedeknsfreqleeeeeeeakhnlekq 1361

Qy     264 -------ITSLRKKSDDPPGNLEPASATN-------ETVSRLVFESPA---PVEMMNPRL 306
                 :  ::|| :|  | || |           | :|:   |  |    :|    ||
```

```
Db   1362 iatlhaqvadmkkkmedsvgcletaeevkrklqkdleglsqrheekvaaydklektktrl 1421

Qy    307 HQPPFGDEIDLN----------TTFDVNTPPTQTSGSQH----------------- 335
Db   1422 qqelddllvdldhqrqsacnlekkqkkfdqllaeektisakyaeerdraeaeareketka 1481

Qy    336 ---------CLPKKLCLERA----RSPMQNVL-------KKVHKVSKPESQLSLGGQRC- 374
Db   1482 lslaraleeameqkaelerlnkqfrtemedlmsskddvgksvhelekskraieqqveemk 1541

Qy    375 --VGELDEELAGAFPLFIRNAVLGQKQPNRTTAESRSSTDVVRIGFDGLGGRTKFIQPRD 432
Db   1542 tqleeledelqatedaklrlev------nlqamkaqferd--------lqgrdeqseekk 1587

Qy    433 TTIIRPV-PVKSKAKSKQKVRIKTVSSASQPKLD 465
Db   1588 kqlvrqvremeaelederkqrsmavaarkklemd 1621


RESULT  34
ABG27218
ID   ABG27218 standard; Protein; 2918 AA.
XX
AC   ABG27218;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #27209.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS91405.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 57577; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   2918 AA;


  Query Match             7.1%;  Score 170.5;  DB 22;  Length 2918;
  Best Local Similarity   18.0%;  Pred. No. 7.6e-05;
  Matches   93;  Conservative  110;  Mismatches  210;  Indels  103;  Gaps   15;

Qy     35 IQWFETAPSRTCPQCRIQVGKKTIINKLFFDLAQEEENVLDAEF---------------- 78
Db   1561 lqkanssatetinklkvqeqeltrl-ridyervsqertvkdqditrfqnslkelqlqkqk 1619

Qy     79 LKNELDSVKAQLSQKDREKRDSQAIIDTLRDTLEERNATVESLQNALNKAEMLC----ST 134
Db   1620 veeelnrlkrtasedsckrkkleeelegmrrslkeqaikitnltqqleqasivkkrsedd 1679

Qy    135 LKKQMKFLEQRQDETKQAREEAHRLKCKMKTMEQIELLLQSQRS-------------EVE 181
Db   1680 lrqqrdvldghlrekqrtqeelrrlssevealrr--qllqeqesvkqahlrnehfqkaie 1737

Qy    182 EMIRDMGVGQSAVEQLAVYCVSLKKEYENLKE-----------ARKATGELADRLKKDLV 230
Db   1738 dksrslneskieierlqsltenltkehlmleeelrnlrleyddlrrgrseadsdknatil 1797

Qy    231 SSRSKLKTLNTELDQAKLELRSAQKDLQSADQEITSLRKKSDDPPGNLEPA-----SATN 285
Db   1798 elrsqlqisnnrtlelqglindlqrerenlrqeiekfqkqaleasnriqesknqctqvvq 1857

Qy    286 ETVSRLV-----FESPAPVEMMNPRLHQPPFGDEIDLNT-----------TFDVNTPPTQ 329
Db   1858 eresllvkikvleqdkarlqrledelnrakstleaetrvkqrlecekqqiqndlnqwktq 1917

Qy    330 TSGSQHCLPKKLCLERARSP-----------MQNVLKKVHK---------VSKPESQLS 368
Db   1918 ysrkeeai-rkieserekserekns lrseierlqaeikrieercrrkledstretqsqle 1976

Qy    369 LGGQRCVGELDEELAGAFPLFIRNAVLGQKQPNRTTAESRSSTDVVRIGFDGLGGRTKFI 428
Db   1977 tersryqreidk---------lrqrpygshretqt--ecewtvdtsklvfdglrkkvtam 2025

Qy    429 QPRDTTIIRPVPVKSKAKSKQKVRIKTVSSASQPKL 464
Db   2026 qlyecqlidkttldkllkgkksv--eevaseiqpfl 2059


RESULT  35
AAR42818
ID   AAR42818 standard; Protein; 1093 AA.
XX
AC   AAR42818;
XX
DT   27-APR-1994  (first entry)
XX
DE   TMF.
XX
KW   TATA modulating factor; TMF; transcription; TATA box; promoter; HIV-1;
KW   human immunodeficiency virus-1; short arm; human chromosome 3; p12-p21;
KW   translocation; cancer.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          437..850
FT                   /label=  TATA binding region
FT   Region          769..777
FT                   /note= "Ubiquitin-mediated protein degradation
FT                        consensus sequence homology region"
FT   Region          454..614
```

```
FT                   /note= "Region with Leucine zipper secondary
FT                         structure"
FT   Region          986..1069
FT                   /note=  "Region with Leucine zipper secondary
FT                         structure"
FT   Region          1070..1078
FT                   /note= "Ubiquitin-mediated protein degradation
FT                         consensus sequence homology region"
XX
PN   WO9320106-A.
XX
PD   14-OCT-1993.
XX
PF   31-MAR-1993;   93WO-US03077.
XX
PR   02-APR-1992;   92US-0862025.
XX
PA   (TEXA ) UNIV TEXAS SYSTEM.
XX
PI   Gaynor RB,  Wu F;
XX
DR   WPI; 1993-336836/42.
DR   N-PSDB; AAQ49397.
XX
PT   New protein cellular factor - capable of binding double stranded
PT   HIV-1 tata region and activating gene expression of HIV-LTR
XX
PS   Claim 2; Fig 1; 75pp; English.
XX
CC   This sequence represents TATA modulating factor (TMF).  TMF is a
CC   protein of mol. wt. 123-130 kD which activates transcription in most
CC   genes, esp. in human immunodeficiency virus-1 (HIV-1) by binding to
CC   the TATA box region of the promoter.  TMF is encoded by the short
CC   arm of human chromosome 3 in the region p12-p21 which is often
CC   involved in translocations in patients having lung and other types
CC   of cancer.
XX
SQ   Sequence   1093 AA;
```

```
  Query Match             7.1%;  Score 170;  DB 14;  Length 1093;
  Best Local Similarity   20.2%;  Pred. No. 2e-05;
  Matches  107;  Conservative   79;  Mismatches  203;  Indels  140;  Gaps   17;

Qy      66 LAQEEENVLDAEFLKNELDSVKAQLSQKDREKRDSQAIIDTLRDTL---EERNATVESLQ 122
           |:::|:     |||  :|:  :|||    :||    :   | |:| :   :| :::: ||:
Db     441 lsekedvcktveflneklekreaqllslskekalleeafdnlkdemfrvkeesssisslk 500

Qy     123 N--------ALNKAEMLC---STLKKQMK----------------FLEQRQDETKQARE 154
           :        |  | :: |      ||::|                |::: :: :   |
Db     501 deftqriaeaekkvqlackerdaakkeiknikeelatrlnssetadllkekdeqirglme 560

Qy     155 EAHRLKCKMKTMEQIELLLQSQRSEVEEMIRDMGVGQSAVEQLAVYCVSLKKEYENLKEA 214
           |  :|  :      |   |:::  | | |:  :      |::|      ||:  :  :|
Db     561 egeklskqqlhnsniikklrakdkenenmvakl---nkkvkeleeelqhlkqvldgkeev 617

Qy     215 RKATGELADRLKKDLVSSRSKLKTLNTELDQAKLELRSAQKDLQSADQEITSLRKKSDDP 274
           |   |   :|   :      |  |  ::|: : : || |  | || :|:| | | :
Db     618 ekqhrenikklnsmverqekdlgrlqvdmdeleeknrsiqaaldsaykeltdlhkanaak 677

Qy     275 PGNLEPASATNETVSRLVFESPAPVEMMNPRLHQP------PFGD--------------- 313
           : |: : |  ::   |  | :|       |         ||
Db     678 dseaqeaalsremkak--eelsaalekaqeearqqqetlaiqvgdlrlalqrteqaaark 735

Qy     314 ------EI--------------DLNTTFDVNTPP--------TQTSGSQHCLPKKL--- 341
                ||               :|: :    | |          | |||    :||
Db     736 edylrheigelqqrlqeaenrnqelsqsvssttrpllrqienlqatlgsqtssweklekn 795

Qy     342 --------------CLERARSPMQNVLKKVHKVSKPESQLSLGGQ---RCVGELDEELAG 384
                        :|| |:  : :|    ::|  ||| ||  |    |   :|: |
Db     796 lsdrlgesqtllaaavereraateellankiqmssmesqnsllrqensrfqaqlese--- 852

Qy     385 AFPLFIRNAVLGQKQPN--------------RTTAESRSSTDVVRIGFDGLGGRTKFIQ 429
```

```
               :| :   :  |              ||  |:|    ::    :    | |  |
Db     853 ------knrlckledennryqvelenlkdeyvrtleetrkektllnsqle--mermkveq 904

Qy     430 PRDTTIIRPVPVKSKAKSKQKV-------RIKTVSSASQPKLDT-FLCQ 470
           |   |     :| | :    |       |  ::|      | | || |
Db     905 erkkaiftqetikekerkpfsvsstptmsrsssisgvdmaglqtsflsq 953
```

```
RESULT  36
ABG06301
ID   ABG06301 standard; Protein; 2017 AA.
XX
AC   ABG06301;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #6292.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS70488.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 36660; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   2017 AA;
```

```
  Query Match             7.1%;  Score 170;  DB 22;  Length 2017;
  Best Local Similarity   22.8%;  Pred. No. 4.9e-05;
```

```
  Matches  102;  Conservative   70;  Mismatches  158;  Indels  118;  Gaps   19;

Qy      59 INKLFFDLAQEEEN-VLDAEFLKNELDSVKAQLSQKDREKRDSQAIIDTL---------- 107
           : |: ||:  | |  ||  | ::::|:   |  :|   ||     |::||
Db     736 lkkvafdvkmeneklvlacedvrhqleeclagnnqlslek---ntivetlkmekgeieae 792

Qy     108 -----RDTLEERN---ATVESLQNALN----------------KAEMLCSTLKKQMKF 141
                :  ||| |    |:| | || |                | :|  : ||| ::
Db     793 lcwakkrlleeankyektieelsnarnlntsalqlehehliklnqkkdmeiaelkknieq 852

Qy     142 LEQRQDETK----QAREEAHRL-----------KCKMKTME-QIELLLQSQRSEVEEMI 184
           ::    |||     : ||  :|            | | :: : | ||   ||     |::
Db     853 mdtdhketkdvlsssleeqkqltqlinkkeifieklkersklqeeldkysqalrkneil 912

Qy     185 R------DMGVGQSAVEQLAVYCVSLKKEYENLKEARKATGELADRLKKDLVSS-RSKLK 237
           |      |  :|    |       |::| | |:| :  |  :||    | |:   |::
Db     913 rqtieekdrslgsmkeen-----nhlqeelerlreeqsrtapvadpktldsvtelasevs 967

Qy     238 TLNTELDQAKLELRSAQKDLQSADQEITSL-------RKKSDDPPGNLEPASATNETVSR 290
            |||  :  : |::  || ::  :|    |       :|: |:     |  :||:   ::
Db     968 qlntikehleeeikhhqkiiedqnqskmqllqslqeqkkemdefryqheqmnath---tq 1024

Qy     291 LVFESPAPVEMMNPRLHQPPFGDEIDLNTTFDVNTPPTQTSGS---QHCLPKKLCLERAR 347
           |  |    :: :   : |        : |         ||  |   |    : | :|
Db    1025 lflekdeeikslqktieq--------iktqlheerqdiqtdnsdifqetkvqslnieng- 1075

Qy     348 SPMQNVLKKVHKVSKPESQLSLGGQRCVGELDEELAGAFPLFIRNAVLGQKQPNRTTAES 407
                : | :|| |::  : |   : | : |:           :| :|  : |
Db    1076 -------sekhdlskaeterlvkg---ikerelei---------kllneknisltkqid 1115

Qy     408 RSSTDVVRIGFDGLGGRTKFIQPRDTTI 435
           : | | |       |  |: || :|  |
Db    1116 qlskdev-------gkltqiiqqkdlei 1136


RESULT  37
AAW63043
ID   AAW63043 standard; Protein; 561 AA.
XX
AC   AAW63043;
XX
DT   26-OCT-1998  (first entry)
XX
DE   Streptococcus uberis bovine lactoferrin binding protein.
XX
KW   Bovine lactoferrin binding protein; LBP; mastitis; vaccine;
KW   diagnosis.
XX
OS   Streptococcus uberis strain su-1 (ATCC 9927).
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..51
FT                   /label= Sig_peptide
FT                   /note= "alternative translation start site at
FT                    Met-11"
FT   Protein         52..561
FT                   /label= Mat_protein
FT   Region          148..199
FT                   /note= "central repeated amino acid sequence A1"
FT   Region          200..212
FT                   /note= "central repeated amino acid sequence B1"
FT   Region          213..271
FT                   /note= "central repeated amino acid sequence C1"
FT   Region          282..325
FT                   /note= "central repeated amino acid sequence A2"
FT   Region          326..339
FT                   /note= "central repeated amino acid sequence B2"
FT   Region          340..397
FT                   /note= "central repeated amino acid sequence C2"
FT   Peptide         525..530
FT                   /note= "surface anchor motif"
XX
PN   WO9821231-A2.
XX
PD   22-MAY-1998.
XX
PF   14-NOV-1997;   97WO-CA00867.
XX
PR   14-NOV-1996;   96US-0031117.
XX
PA   (UYSA-) UNIV SASKATCHEWAN.
XX
PI   Jiang M,  Maclachlan PR,  Potter AA;
XX
DR   WPI; 1998-297860/26.
DR   N-PSDB; AAV42601.
XX
PT   Immunogenic Streptococcus uberis protein(s) that bind bovine
PT   lactoferrin - associated regulatory protein, useful in vaccines for
PT   treatment and prevention of mastitis
XX
PS   Claim 2; Fig 2A-C; 105pp; English.
XX
CC   This is the bovine lactoferrin binding protein (LBP) of
CC   Streptococcus uberis su-1.  Its amino acid sequence was deduced
CC   from the novel isolated lbp gene (see AAV42601).   The LBP is
CC   lactoferrin species-specific; human lactoferrin does not
CC   effectively block binding of bovine lactoferrin.  The invention
CC   provides recombinant vectors, transformed host cells and methods of
CC   producing recombinant bovine LBP of S. uberis.  The bovine LBP,
CC   immunogenic fragments and/or chimeric proteins can be used, either
CC   alone or in combination with other antigens, in novel subunit
CC   vaccines for the prevention and treatment of S. uberis infections,
CC   particularly mastitis, as well as in diagnostic methods for
CC   determining the presence of S. uberis infections.
XX
SQ   Sequence   561 AA;


  Query Match             7.1%;  Score 169.5;  DB 19;  Length 561;
  Best Local Similarity   23.6%;  Pred. No. 8.6e-06;
  Matches   63;  Conservative   46;  Mismatches  101;  Indels   57;  Gaps    8;

Qy      56 KTIINKLFFDLAQEEENVLDAEFLKN------ELDSVKAQLSQKDREKRDSQAIIDTLRD 109
           | :  ||  | :::|   |  || ::      ||      | : ::   | :|    |::
Db     243 kdltekl--dasrkehealakefaesqkgyekeladkhtalgeaekrnadleagnkelke 300

Qy     110 TLEERNATVESLQNALNKAEMLCSTLKKQMKFLEQRQDETKQAREEAHRLKCKMKTMEQI 169
            ||      : ||  : |||       ::|| |  : :|   |:||    | |:   |:
Db     301 nlemaegisddlqkkvmkae-------qemkelsaqlee---akeeletekaklaeseke 350

Qy     170 ELLLQSQR-------------SEVEEMIRDMGVGQSAVEQLAVYCVSLKKEYENLKEAR 215
             |  :|               :||::: ::   :   |:|      |:|::| :|  :
Db     351 naklteerdaakkeaekvpeleeqveklveeitaakkeaeelqakaeglekdfeavkaek 410

Qy     216 KATGELADRLKKD-----------LVSSRSKLKTLNTELDQAKLELRSAQKDLQSADQEI 264
           :|      :||:|           |      || |  :||:|| |   | |: | : :|
Db     411 ealeaeiaklkedhqkevdalnalladkekmlknlqdqldkakee---amkneqmsqeek 467

Qy     265 TSLRKKSD-----------DPPGNLEP 280
             |: : |           | |  : |
Db     468 aklqaeldqakkelaekikdmpnkvap 494


RESULT  38
AAY59270
ID   AAY59270 standard; Protein; 1090 AA.
XX
AC   AAY59270;
XX
DT   17-APR-2000  (first entry)
XX
DE   Human huntingtin-interacting protein (HIP1).
XX
KW   Huntingtin Interacting Protein; HIP; death effector domain; DED; human;
KW   apotosis; HIP-apoptosis modulating protein; cell death; gene therapy;
```

```
KW   Huntington's disease; nootropic; anticonvulsant; cytostatic;
KW   chromosome 7q11.23.
XX
OS   Homo sapiens.
XX
PN   WO9960986-A2.
XX
PD   02-DEC-1999.
XX
PF   27-MAY-1999;   99WO-US11743.
XX
PR   27-MAY-1998;   98US-0085199.
XX
PA   (UYBR-) UNIV BRITISH COLUMBIA.
PA   (MERI ) MERCK FROSST CANADA INC.
XX
PI   Kalchman M,  Hayden MR,  Hackam A,  Chopra V,  Nicholson DW;
PI   Vallaincourt JP,  Rasper DM;
XX
DR   WPI; 2000-097055/08.
XX
PT   Novel proteins useful for treating Huntington's disease by gene therapy
PT   techniques, and cancers -
XX
PS   Claim 1; Page 57-62; 91pp; English.
XX
CC   The invention relates to Huntingtin Interacting Protein (HIP), that
CC   includes a death effector domain (DED), suggesting apototic function.
CC   Proteins with DED (referred as HIP-apoptosis modulating proteins) are
CC   useful for inducing apoptotic death in cells. The HIP is a logical target
CC   for therapy in Huntington's disease since it has a differential apoptotic
CC   activity, modulated by interaction with Huntingtin having normal and
CC   expanded  repeats. HIP is also used as a therapeutic agent to introduce
CC   apoptosis in cancer cells. Increasing expression of normal (non-expanded)
CC   Huntingtin or the HIP-apoptotic modulator-binding  portion, a modified
CC   HIP-apoptotic modulator in which the DED has been deleted, is helpful for
CC   treating Huntington's disease. The present sequence represents the human
CC   HIP1.
XX
SQ   Sequence   1090 AA;


  Query Match             7.1%;  Score 169;  DB 21;  Length 1090;
  Best Local Similarity   23.0%;  Pred. No. 2.5e-05;
  Matches  110;  Conservative   94;  Mismatches  192;  Indels   82;  Gaps   23;

Qy      50 RIQVGKKTIINKLFFDLAQE----EENVLDAEFLKNELDSVKAQLSQKDREKRDSQAIID 105
           |: :  |  :::|  |||::    ::   | |||: ||| :: |    :: :| | : |:
Db     453 rvvlqlkghvseleadlaeqqhlrqqaaddceflraeldelrrqredtekaqr-slseie 511

Qy     106 TLRDTLEERNATVESLQNAL--NKAEML--CSTLKKQMKFLEQRQ----DETKQAREEAH 157
             |:| : ::   : |  | |::|   : : ||:   | |     | |:  :
Db     512 rkaqaneqrysklkekyselvqnhadllrknaevtkqvsmarqaqvdlerekkeledsle 571

Qy     158 RL--KCKMKTMEQIELLLQSQRSEVEEMIRDMGVGQSAVE-------QLAVYCVSLKKEY 208
           |:  : : || ||:| :|:| : |:    |:: | | ::|         |     |:||
Db     572 risdqgqrktqeqle-vleslkqelatsqrelqvlqgsletsaqseanwaaefaeleker 630

Qy     209 ENLKEARKATGELADRLKKDLVSSRSKLKTLNTEL-----DQAKLELRSAQKDLQSADQE 263
           ::|       |    |:|:|  :: || :    :     || |: |  ::|   :|:|
Db     631 dslvsgaahreeelsalrkelqdtqlklasteesmcqlakdqrkmllvgsrk---aaeqv 687

Qy     264 ITSLRKKSDDPP-----GNLEPASATNETVSRLVFESP-------APVEMMNPRLHQPPF 311
           |     : ::||     |: :   :|  ::|  : :          |  | ::  ||
Db     688 iqdalnqleeppliscagsadhllstvtsisscieqlekswsqylacpedisgllhsitl 747

Qy     312 -----GDEIDLNTTFDVNTPPTQT-SGSQHC------------LPKKLCLERARS-PMQ 351
                | |    |  :  ||    | :: |             | ::  || | |  |:
Db     748 lahltsdaiahgattclrappepadslteackqygretlaylasleeegslenadstamr 807

Qy     352 NVLKKVHKVSKPESQLSLG---GQRCVGEL-DEELAGAFPLFIRNAVLGQKQPNRTTAES 407
           | | |:  :   |  |  |    |  :|:| |:|:| |    |  |     :     ::|
Db     808 nclskikaig--eellprgldikqeelgdlvdkema-atsaaietat---arieemlsks 861
```

```
Qy     408 RSSTDVVRIGFDG--LGGRTKFIQPRDTTIIRPVPVKSKAKSKQKVRIKTVSSASQPK 463
           |:    |:: :   ||  |  :|     |:        :|  |:  :::    : ||
Db     862 ragdtgvklevnerilgcctslmqaiqvliv-------askdlqreivesgrgtaspk 912


RESULT  39
AAU31067
ID   AAU31067 standard; Protein; 560 AA.
XX
AC   AAU31067;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Novel human secreted protein #1558.
XX
KW   Human; vaccination; gene therapy; nutritional supplement;
KW   stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW   immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200179449-A2.
XX
PD   25-OCT-2001.
XX
PF   16-APR-2001; 2001WO-US08656.
XX
PR   18-APR-2000; 2000US-0552929.
PR   26-JAN-2001; 2001US-0770160.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-611725/70.
XX
PT   Nucleic acids encoding a range of human polypeptides, useful in genetic
PT   vaccination, testing and therapy -
XX
PS   Claim 20; Page 399; 765pp; English.
XX
CC   The invention relates to novel human secreted polypeptides. The
CC   polypeptides and antibodies to the polypeptides are useful for
CC   determining the presence of or predisposition to a disease associated
CC   with altered levels of polypeptide. The polypeptides are also useful for
CC   identifying agents (agonists and antagonists) that bind to them. Cells
CC   expressing the proteins are useful for identifying a therapeutic agent
CC   for use in treatment of a pathology related to aberrant expression or
CC   physiological interactions of the polypeptide. Vectors comprising
CC   the nucleic acids encoding the polypeptides and cells genetically
CC   engineered to express them are also useful for producing the proteins.
CC   The proteins are useful in genetic vaccination, testing and
CC   therapy, and can be used as nutritional supplements. They may be used to
CC   increase stem cell proliferation; to regulate haematopoiesis; and in
CC   bone, cartilage, tendon  and/or nerve tissue growth or regeneration;
CC   immune suppression and/or stimulation; as anti-inflammatory agents; and
CC   in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
CC   sequences of novel human secreted proteins of the invention.
XX
SQ   Sequence   560 AA;


  Query Match             7.0%;  Score 168.5;  DB 22;  Length 560;
  Best Local Similarity   23.8%;  Pred. No. 1e-05;
  Matches   65;  Conservative   54;  Mismatches   99;  Indels   55;  Gaps   10;

Qy      65 DLAQEEENVLDAEFLKNELDSVKAQLSQK----------DREKRDSQAIIDTLRDTLEE 113
           || | :|::  ||  : : |  | :|:::             |||  :| |  | : |||
Db     309 dlmqlqedlaaaerarkqadlekeelaeelasslsgrnalqdekrrleariaqleeelee 368

Qy     114 RNATVESLQNALNK----AEMLC-------STLKKQMKFLEQRQDETKQAREEAHRL--- 159
             :|:: : : |    || |        || :|     :| : : |: | : | :
```

```
Db    369 eqgnmeamsdrvrkatqqaeqlsnelaterstaqknesarqqlerqnkelrsklhemega 428

Qy    160 -KCKMK-TMEQIELLLQSQRSEVEEMIRDMGVGQSAVEQLAVYCVSLKKEYENLKEARKA 217
           | | | |:  :|  :     :||:  |:    |:| : |      ||:    ::: ||
Db    429 vkskfkstiaaleakiaqleeqveqearek---qaatkslkqkdkklkeillqvederk- 484

Qy    218 TGELADRLKKDLVSSRSKLKTLNTELDQAKLELRSAQKDLQSADQEITSLRKKSDDPPGN 277
             :|:: |:      :::| |  :|::|: |            | | : |:|
Db    485 ---maeqykeqaekgnarvkqlkrqleeaeee----------sqrinanrrklq---re 527

Qy    278 LEPASATNETVSRLVFESPAPVEMMNPRLHQPP 310
          |: |: :|| : |        |  :  :|  ||
Db    528 ldeatesneamgr-------evnalksklrgpp 553


RESULT  40
AAR66929
ID   AAR66929 standard; Protein; 576 AA.
XX
AC   AAR66929;
XX
DT   01-SEP-1995  (first entry)
XX
DE   AMML chromosome inv(16) product.
XX
KW   AMML; acute myelomonocytic leukemia; chromosome-16; inversion;
KW   inv(16); CBF-beta; CBFB gene; transcription factor; myosin; MYH11;
KW   SMMHC.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..164
FT                   /label= CBFB
FT   Peptide         165..576
FT                   /label= MYH11
XX
PN   WO9504067-A.
XX
PD   09-FEB-1995.
XX
PF   26-JUL-1994;   94WO-US08530.
XX
PR   29-JUL-1993;   93US-0099869.
XX
PA   (UNMI ) UNIV MICHIGAN.
PA   (TEXA ) UNIV TEXAS SYSTEM.
XX
PI   Claxton D,  Collins FS,  Liu P,  Siciliano MJ;
XX
DR   WPI; 1995-082178/11.
DR   N-PSDB; AAQ84588.
XX
PT   Novel DNA spanning the pericentric inversion of chromosome 16 -
PT   for the screening of acute myeloid leukaemia
XX
PS   Claim 4; Page 28-30; 78pp; English.
XX
CC   PCR was performed on total cellular RNA from 5 AMML patients having
CC   a pericentric inversion of chromosome-16, M4Eo subtype.  Sequencing
CC   showed the inv(16) fusion to comprise a sequence from the CBFB
CC   gene, encoding a novel transcription factor, and the MYH11 gene,
CC   encoding smooth muscle myosin heavy chain.  In 3 patients, nt 1-492
CC   of the CBFB gene were fused to nt 1921 of MYH11 (shown in
CC   AAQ84588; predicted aa sequence in AAR66929).  Probes based on inv(16)
CC   can be used for diagnosis of AMML.
XX
SQ   Sequence   576 AA;


  Query Match             7.0%;  Score 168.5;  DB 16;  Length 576;
  Best Local Similarity   23.8%;  Pred. No. 1.1e-05;
  Matches   65;  Conservative   54;  Mismatches   99;  Indels   55;  Gaps   10;

Qy     65 DLAQEEENVLDAEFLKNELDSVKAQLSQK----------DREKRDSQAIIDTLRDTLEE 113
          || | :|::  ||  : : |  | :|:::            |||  :| |  | : |||
Db    326 dlmqlqedlaaaerarkqadlekeelaeelasslsgrnalqdekrrleariaqleeelee 385

Qy    114 RNATVESLQNALNK----AEMLC-------STLKKQMKFLEQRQDETKQAREEAHRL--- 159
             :|:: : : |    || |        || :|     :| : : |: | : | :
Db    386 eqgnmeamsdrvrkatqqaeqlsnelaterstaqknesarqqlerqnkelrsklhemega 445

Qy    160 -KCKMK-TMEQIELLLQSQRSEVEEMIRDMGVGQSAVEQLAVYCVSLKKEYENLKEARKA 217
           | | | |:  :|  :     :||:  |:    |:| : |      ||:    ::: ||
Db    446 vkskfkstiaaleakiaqleeqveqearek---qaatkslkqkdkklkeillqvederk- 501

Qy    218 TGELADRLKKDLVSSRSKLKTLNTELDQAKLELRSAQKDLQSADQEITSLRKKSDDPPGN 277
             :|:: |:      :::| |  :|::|: |            | | : |:|
Db    502 ---maeqykeqaekgnarvkqlkrqleeaeee----------sqrinanrrklq---re 544

Qy    278 LEPASATNETVSRLVFESPAPVEMMNPRLHQPP 310
          |: |: :|| : |        |  :  :|  ||
Db    545 ldeatesneamgr-------evnalksklrgpp 570


Search completed: September  4, 2002, 16:09:06
Job time: 8130 sec
```